

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

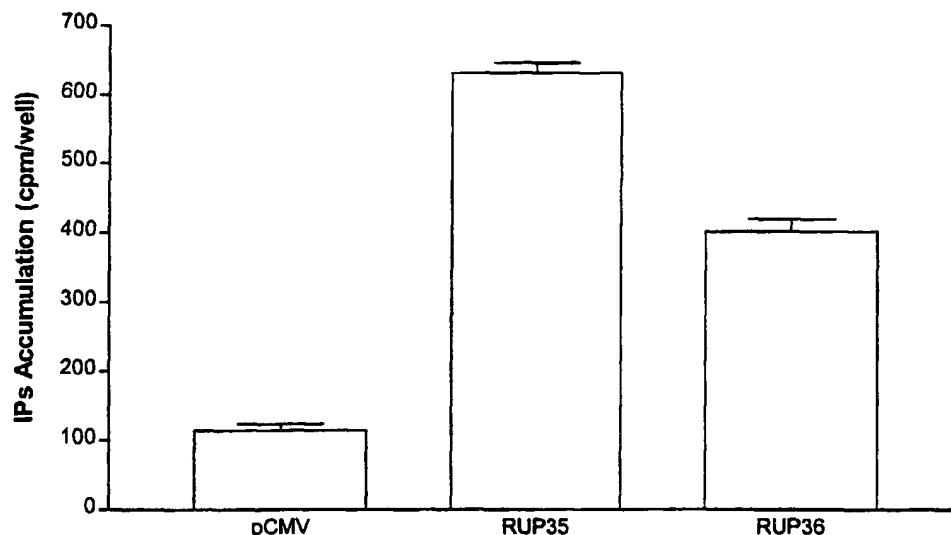
(10) International Publication Number
WO 02/42461 A2

- (51) International Patent Classification⁷: C12N 15/12, C07K 14/705
- (21) International Application Number: PCT/US01/44386
- (22) International Filing Date:
26 November 2001 (26.11.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- | | | |
|------------|-------------------------------|----|
| 60/253,404 | 27 November 2000 (27.11.2000) | US |
| 60/255,366 | 12 December 2000 (12.12.2000) | US |
| 60/270,266 | 20 February 2001 (20.02.2001) | US |
| 60/270,286 | 20 February 2001 (20.02.2001) | US |
| 60/282,032 | 6 April 2001 (06.04.2001) | US |
| 60/282,356 | 6 April 2001 (06.04.2001) | US |
| 60/282,358 | 6 April 2001 (06.04.2001) | US |
| 60/282,365 | 6 April 2001 (06.04.2001) | US |
| 60/290,917 | 14 May 2001 (14.05.2001) | US |
| 60/309,208 | 31 July 2001 (31.07.2001) | US |
- (71) Applicant (for all designated States except US): ARENA PHARMACEUTICALS, INC. [US/US]; 6166 Nancy Drive, San Diego, CA 92121 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): CHEN, Ruoping [CN/US]; 5296 Timber Branch Way, San Diego, CA 92150 (US). CHU, Zhi, Liang [CN/US]; 11489 Creekstone Lane, San Diego, CA 92128 (US). DANG, Huong, T. [VN/US]; 5352 Oak Park Drive, San Diego, CA 92105 (US). LOWITZ, Kevin, P. [US/US]; 8031 Caminito de Pizza, Apt. C, San Diego, CA 92108 (US). PRIDE, Cameron [US/US]; 11026 West Ocean Air Drive #1128, San Diego, CA 92130 (US).
- (74) Agents: STRAHER, Michael, P. et al.; Woodcock Washburn LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

[Continued on next page]

(54) Title: ENDOGENOUS AND NON-ENDOGENOUS VERSIONS OF HUMAN G PROTEIN-COUPLED RECEPTORS

IP3 Assay in 293 Cells



(57) Abstract: The invention disclosed in this patent document relates to transmembrane receptors, more particularly to a human G protein-coupled receptor for which the endogenous ligand is unknown, and to mutated (non-endogenous) versions of the human GPCRs for evidence of constitutive activity.

WO 02/42461 A2

MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv)) for US only

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

**ENDOGENOUS AND NON-ENDOGENOUS VERSIONS OF
HUMAN G PROTEIN-COUPLED RECEPTORS**

5

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Serial Number 09/170,496, filed with the United States Patent and Trademark Office on October 13, 1998 and its
10 corresponding PCT application number PCT/US99/23938, published as WO 00/22129 on April 20, 2000. This document claims the benefit of priority from the following provisional applications, all filed via U.S. Express Mail with the United States Patent and Trademark Office on the indicated dates: U.S. Provisional Number 60/253,404, filed November 27, 2000; U.S. Provisional Number 60/255,366, filed December 12, 2000;
15 U.S. Provisional Number 60/270,286 filed February 20, 2001; U.S. Provisional Number 60/282,356, filed April 6, 2001, which claims priority from U.S. Provisional Number 60/270,266, filed February 20, 2001; U.S. Provisional Number 60/282,032, filed April 6, 2001; U.S. Provisional Number 60/282,358, filed April 6, 2001; U. S. Provisional Number 60/282,365, filed April 6, 2001; U.S. Provisional Number 60/290,917, filed
20 May 14, 2001; U.S. Provisional Number 60/309,208, filed July 31, 2001; the disclosures of which are incorporated in their entirety by reference.

FIELD OF THE INVENTION

25 The present invention relates to transmembrane receptors, in some embodiments to G protein-coupled receptors and, in some preferred embodiments, to endogenous GPCRs that are altered to establish or enhance constitutive activity of the receptor. In some embodiments, the constitutively activated GPCRs will be used for the direct identification of candidate compounds as receptor agonists or inverse agonists having applicability as
30 therapeutic agents.

BACKGROUND OF THE INVENTION

Although a number of receptor classes exist in humans, by far the most abundant and therapeutically relevant is represented by the G protein-coupled receptor (GPCR) class.

5 It is estimated that there are some 30,000-40,000 genes within the human genome, and of these, approximately 2% are estimated to code for GPCRs. Receptors, including GPCRs, for which the endogenous ligand has been identified, are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified are referred to as "orphan" receptors.

10 GPCRs represent an important area for the development of pharmaceutical products: from approximately 20 of the 100 known GPCRs, approximately 60% of all prescription pharmaceuticals have been developed. For example, in 1999, of the top 100 brand name prescription drugs, the following drugs interact with GPCRs (diseases and/or disorders treated are indicated in parentheses):

15	Claritin® (allergies)	Prozac® (depression)	Vasotec® (hypertension)
	Paxil® (depression)	Zoloft® (depression)	Zyprexa® (psychotic disorder)
	Cozaar® (hypertension)	Imitrex® (migraine)	Zantac® (reflux)
	Propulsid® (reflux disease)	Risperdal® (schizophrenia)	Serevent® (asthma)
	Pepcid® (reflux)	Gaster® (ulcers)	Atrovent® (bronchospasm)
20	Effexor® (depression)	Depakote® (epilepsy)	Cardura® (prostatic hypertrophy)
	Allegra® (allergies)	Lupron® (prostate cancer)	Zoladex® (prostate cancer)
	Diprivan® (anesthesia)	BuSpar® (anxiety)	Ventolin® (bronchospasm)
	Hytrin® (hypertension)	Wellbutrin® (depression)	Zyrtec® (rhinitis)
	Plavix® (MI/stroke)	Toprol-XL® (hypertension)	Tenormin® (angina)
25	Xalatan® (glaucoma)	Singulair® (asthma)	Diovan® (hypertension)
	Harnal® (prostatic hyperplasia)		

(Med Ad News 1999 Data).

GPCRs share a common structural motif, having seven sequences of between 22 to
 30 24 hydrophobic amino acids that form seven alpha helices, each of which spans the membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmembrane-2 (TM-2), *etc.*). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or

“extracellular” side, of the cell membrane (these are referred to as “extracellular” regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and transmembrane-5 and transmembrane-6 on the interior, or “intracellular” side, of the cell membrane (these are referred to as “intracellular” regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The “carboxy” (“C”) terminus of the receptor lies in the intracellular space within the cell, and the “amino” (“N”) terminus of the receptor lies in the extracellular space outside of the cell.

Generally, when an endogenous ligand binds with the receptor (often referred to as “activation” of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular “G-protein.” It has been reported that GPCRs are “promiscuous” with respect to G proteins, *i.e.*, that a GPCR can interact with more than one G protein. *See*, Kenakin, T., 43 *Life Sciences* 1095 (1988). Although other G proteins exist, currently, G_q, G_s, G_i, G_z and G_o are G proteins that have been identified. Ligand-activated GPCR coupling with the G-protein initiates a signaling cascade process (referred to as “signal transduction”). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. Although not wishing to be bound to theory, it is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein.

Under physiological conditions, GPCRs exist in the cell membrane in equilibrium between two different conformations: an “inactive” state and an “active” state. A receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to initiate signal transduction leading to a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response.

A receptor may be stabilized in an active state by a ligand or a compound such as a drug. Recent discoveries, including but not exclusively limited to modifications to the amino acid sequence of the receptor, provide means other than ligands or drugs to promote and stabilize the receptor in the active state conformation. These means effectively stabilize the receptor in an active state by simulating the effect of a ligand binding to the receptor. Stabilization by such ligand-independent means is termed “constitutive receptor activation.”

SUMMARY OF THE INVENTION

Disclosed herein are endogenous and non-endogenous versions of human GPCRs and uses thereof.

Some embodiments of the present invention relate to a G protein-coupled
5 receptor encoded by an amino acid sequence of SEQ.ID.NO.:2, non-endogenous, constitutively activated versions of the same, and host cells comprising the same..

Some embodiments of the present invention relate to a plasmid comprising a vector and the cDNA of SEQ.ID.NO.:1 and host cells comprising the same.

Some embodiments of the present invention relate to a G protein-coupled
10 receptor encoded by an amino acid sequence of SEQ.ID.NO.:4, non-endogenous, constitutively activated versions of the same, and host cells comprising the same..

Some embodiments of the present invention relate to a plasmid comprising a vector and the cDNA of SEQ.ID.NO.:3, non-endogenous, constitutively activated versions of the same, and host cells comprising the same.

15 Some embodiments of the present invention relate to G protein-coupled receptor encoded by an amino acid sequence of SEQ.ID.NO.:6, non-endogenous, constitutively activated versions of the same, and host cells comprising the same.

Some embodiments of the present invention relate to a plasmid comprising a vector and the cDNA of SEQ.ID.NO.:5 and host cells comprising the same.

20 Some embodiments of the present invention relate to a G protein-coupled receptor encoded by an amino acid sequence of SEQ.ID.NO.:8, non-endogenous, constitutively activated versions of the same, and host cells comprising the same.

Some embodiments of the present invention relate to a plasmid comprising a vector and the cDNA of SEQ.ID.NO.:7, non-endogenous, constitutively activated
25 versions of the same, and host cells comprising the same.

Some embodiments of the present invention relate to a G protein-coupled receptor encoded by an amino acid sequence of SEQ.ID.NO.:10, non-endogenous, constitutively activated versions of the same, and host cells comprising the same.

Some embodiments of the present invention relate to a plasmid comprising a
30 vector and the cDNA of SEQ.ID.NO.:9 and host cells comprising the same.

Some embodiments of the present invention relate to a G protein-coupled receptor encoded by an amino acid sequence of SEQ.ID.NO.:12, non-endogenous, constitutively activated versions of the same, and host cells comprising the same.

5 Some embodiments of the present invention relate to a plasmid comprising a vector and the cDNA of SEQ.ID.NO.:11, and host cells comprising the same.

Some embodiments of the present invention relate to a G protein-coupled receptor encoded by an amino acid sequence of SEQ.ID.NO.:14, constitutively activated versions of the same, and host cells comprising the same.

10 Some embodiments of the present invention relate to a plasmid comprising a vector and the cDNA of SEQ.ID.NO.:13 and host cells comprising the same.

Some embodiments of the present invention relate to a G protein-coupled receptor encoded by an amino acid sequence of SEQ.ID.NO.:16, constitutively activated versions of the same, and host cells comprising the same.

15 Some embodiments of the present invention relate to a plasmid comprising a vector and the cDNA of SEQ.ID.NO.:15 and host cells comprising the same.

Some embodiments of the present invention relate to a G protein-coupled receptor encoded by an amino acid sequence of SEQ.ID.NO.:18, constitutively activated versions of the same, and host cells comprising the same.

20 Some embodiments of the present invention relate to a plasmid comprising a vector and the cDNA of SEQ.ID.NO.:17 and host cells comprising the same.

Some embodiments of the present invention relate to a G protein-coupled receptor encoded by an amino acid sequence of SEQ.ID.NO.:20, constitutively activated versions of the same, and host cells comprising the same.

25 Some embodiments of the present invention relate to a plasmid comprising a vector and the cDNA of SEQ.ID.NO.:19 and host cells comprising the same.

BRIEF DESCRIPTION OF THE DRAWINGS

30 **Figure 1** is a graphic representation of activation of RUP32, $G_q(\text{del})/G_i$, RUP32 co-transfected with $G_q(\text{del})/G_i$, and CMV (control; expression vector) in a second messenger assay measuring the accumulation of inositol phosphate (IP_3) utilizing 293 cells.

Figure 2 provides an illustration of second messenger IP_3 production from endogenous version RUP35 and RUP36 as compared with the control ("CMV").

DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

AGONISTS shall mean materials (*e.g.*, ligands, candidate compounds) that activate the intracellular response when they bind to the receptor, or enhance GTP binding to membranes. In some embodiments, **AGONISTS** are those materials not previously known to activate the intracellular response when they bind to the receptor or to enhance GTP binding to membranes.

AMINO ACID ABBREVIATIONS used herein are set out in Table A:

TABLE A

ALANINE	ALA	A
ARGININE	ARG	R
ASPARAGINE	ASN	N
ASPARTIC ACID	ASP	D
CYSTEINE	CYS	C
GLUTAMIC ACID	GLU	E
GLUTAMINE	GLN	Q
GLYCINE	GLY	G
HISTIDINE	HIS	H
ISOLEUCINE	ILE	I
LEUCINE	LEU	L
LYSINE	LYS	K

METHIONINE	MET	M
PHENYLALANINE	PHE	F
PROLINE	PRO	P
SERINE	SER	S
THREONINE	THR	T
TRYPTOPHAN	TRP	W
TYROSINE	TYR	Y
VALINE	VAL	V

ANTAGONIST shall mean materials (e.g., ligands, candidate compounds) that competitively bind to the receptor at the same site as the agonists but which do not activate the intracellular response initiated by the active form of the receptor, and can thereby inhibit the intracellular responses by agonists. **ANTAGONISTS** do not diminish the baseline intracellular response in the absence of an agonist. In some embodiments, **ANTAGONISTS** are those materials not previously known to activate the intracellular response when they bind to the receptor or to enhance GTP binding to membranes.

CANDIDATE COMPOUND shall mean a molecule (for example, and not limitation, a chemical compound) that is amenable to a screening technique. Preferably, the phrase "candidate compound" does not include compounds which were publicly known to be compounds selected from the group consisting of inverse agonist, agonist or antagonist to a receptor, as previously determined by an indirect identification process ("indirectly identified compound"); more preferably, not including an indirectly identified compound which has previously been determined to have therapeutic efficacy in at least one mammal; and, most preferably, not including an indirectly identified compound which has previously been determined to have therapeutic utility in humans.

COMPOSITION means a material comprising at least one component; a "pharmaceutical composition" is an example of a composition.

COMPOUND EFFICACY shall mean a measurement of the ability of a compound to inhibit or stimulate receptor functionality; i.e. the ability to activate/inhibit a

signal transduction pathway, as opposed to receptor binding affinity. Exemplary means of detecting compound efficacy are disclosed in the Example section of this patent document.

CODON shall mean a grouping of three nucleotides (or equivalents to nucleotides) which generally comprise a nucleoside (adenosine (A), guanosine (G), cytidine (C), uridine (U) and thymidine (T)) coupled to a phosphate group and which, when translated, encodes an amino acid.

CONSTITUTIVELY ACTIVATED RECEPTOR shall mean a receptor subjected to constitutive receptor activation. A constitutively activated receptor can be endogenous or non-endogenous.

CONSTITUTIVE RECEPTOR ACTIVATION shall mean stabilization of a receptor in the active state by means other than binding of the receptor with its ligand or a chemical equivalent thereof.

CONTACT or **CONTACTING** shall mean bringing at least two moieties together, whether in an in vitro system or an in vivo system.

DIRECTLY IDENTIFYING or **DIRECTLY IDENTIFIED**, in relationship to the phrase "candidate compound", shall mean the screening of a candidate compound against a constitutively activated receptor, preferably a constitutively activated orphan receptor, and most preferably against a constitutively activated G protein-coupled cell surface orphan receptor, and assessing the compound efficacy of such compound. This phrase is, under no circumstances, to be interpreted or understood to be encompassed by or to encompass the phrase "indirectly identifying" or "indirectly identified."

ENDOGENOUS shall mean a material that a mammal naturally produces. **ENDOGENOUS** in reference to, for example and not limitation, the term "receptor," shall mean that which is naturally produced by a mammal (for example, and not limitation, a human) or a virus. By contrast, the term **NON-ENDOGENOUS** in this context shall mean that which is not naturally produced by a mammal (for example, and not limitation, a human) or a virus. For example, and not limitation, a receptor which is not constitutively active in its endogenous form, but when manipulated becomes constitutively active, is most preferably referred to herein as a "non-endogenous, constitutively activated receptor." Both terms can be utilized to describe both "in vivo" and "in vitro" systems. For example, and not limitation, in a screening approach, the endogenous or non-endogenous receptor may be

in reference to an in vitro screening system. As a further example and not limitation, where the genome of a mammal has been manipulated to include a non-endogenous constitutively activated receptor, screening of a candidate compound by means of an in vivo system is viable.

5 **G PROTEIN COUPLED RECEPTOR FUSION PROTEIN** and **GPCR FUSION PROTEIN**, in the context of the invention disclosed herein, each mean a non-endogenous protein comprising an endogenous, constitutively activate GPCR or a non-endogenous, constitutively activated GPCR fused to at least one G protein, most preferably the alpha (α) subunit of such G protein (this being the subunit that binds GTP), with the G
10 protein preferably being of the same type as the G protein that naturally couples with endogenous orphan GPCR. For example, and not limitation, in an endogenous state, if the G protein "G_s α " is the predominate G protein that couples with the GPCR, a GPCR Fusion Protein based upon the specific GPCR would be a non-endogenous protein comprising the GPCR fused to G_s α ; in some circumstances, as will be set forth below, a non-predominant
15 G protein can be fused to the GPCR. The G protein can be fused directly to the C-terminus of the constitutively active GPCR or there may be spacers between the two.

HOST CELL shall mean a cell capable of having a Plasmid and/or Vector incorporated therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as a autonomous molecule as the Host Cell replicates (generally, the Plasmid is
20 thereafter isolated for introduction into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. In some embodiments the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

25 **INDIRECTLY IDENTIFYING** or **INDIRECTLY IDENTIFIED** means the traditional approach to the drug discovery process involving identification of an endogenous ligand specific for an endogenous receptor, screening of candidate compounds against the receptor for determination of those which interfere and/or compete with the ligand-receptor interaction, and assessing the efficacy of the compound for affecting at least one second
30 messenger pathway associated with the activated receptor.

INHIBIT or **INHIBITING**, in relationship to the term "response" shall mean that a response is decreased or prevented in the presence of a compound as opposed to in the absence of the compound.

INVERSE AGONISTS shall mean materials (*e.g.*, ligand, candidate compound) which bind to either the endogenous form of the receptor or to the constitutively activated form of the receptor, and which inhibit the baseline intracellular response initiated by the active form of the receptor below the normal base level of activity which is observed in the absence of agonists, or decrease GTP binding to membranes. Preferably, the baseline intracellular response is inhibited in the presence of the inverse agonist by at least 30%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, and most preferably at least 99% as compared with the baseline response in the absence of the inverse agonist.

KNOWN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has been identified.

LIGAND shall mean a molecule specific for a naturally occurring receptor.

MUTANT or **MUTATION** in reference to an endogenous receptor's nucleic acid and/or amino acid sequence shall mean a specified change or changes to such endogenous sequences such that a mutated form of an endogenous, non-constitutively activated receptor evidences constitutive activation of the receptor. In terms of equivalents to specific sequences, a subsequent mutated form of a human receptor is considered to be equivalent to a first mutation of the human receptor if (a) the level of constitutive activation of the subsequent mutated form of a human receptor is substantially the same as that evidenced by the first mutation of the receptor; and (b) the percent sequence (amino acid and/or nucleic acid) homology between the subsequent mutated form of the receptor and the first mutation of the receptor is at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, and most preferably at least 99%. In some embodiments, owing to the fact that some preferred cassettes disclosed herein for achieving constitutive activation include a single amino acid and/or codon change between the endogenous and the non-endogenous forms of the GPCR, it is preferred that the percent sequence homology should be at least 98%.

NON-ORPHAN RECEPTOR shall mean an endogenous naturally occurring molecule specific for an identified ligand wherein the binding of a ligand to a receptor activates an intracellular signaling pathway.

ORPHAN RECEPTOR shall mean an endogenous receptor for which the ligand
5 specific for that receptor has not been identified or is not known.

PHARMACEUTICAL COMPOSITION shall mean a composition comprising at least one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, and not limitation, a human). Those of ordinary skill in the art will understand and appreciate the techniques appropriate
10 for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

PLASMID shall mean the combination of a Vector and cDNA. Generally, a Plasmid is introduced into a Host Cell for the purposes of replication and/or expression of the cDNA as a protein.

SECOND MESSENGER shall mean an intracellular response produced as a result
15 of receptor activation. A second messenger can include, for example, inositol triphosphate (IP₃), diacylglycerol (DAG), cyclic AMP (cAMP), and cyclic GMP (cGMP). Second messenger response can be measured for a determination of receptor activation. In addition, second messenger response can be measured for the direct identification of candidate
20 compounds, including for example, inverse agonists, agonists, and antagonists.

SIGNAL TO NOISE RATIO shall mean the signal generated in response to activation, amplification, or stimulation wherein the signal is above the background noise or the basal level in response to non-activation, non-amplification, or non-stimulation.

SPACER shall mean a translated number of amino acids that are located after the
25 last codon or last amino acid of a gene, for example a GPCR of interest, but before the start codon or beginning regions of the G protein of interest, wherein the translated number amino acids are placed in-frame with the beginnings regions of the G protein of interest. The number of translated amino acids can be tailored according to the needs of the skilled artisan and is generally from about one amino acid, preferably two amino acids, more
30 preferably three amino acids, more preferably four amino acids, more preferably five amino acids, more preferably six amino acids, more preferably seven amino acids, more preferably

eight amino acids, more preferably nine amino acids, more preferably ten amino acids, more preferably eleven amino acids, and even more preferably twelve amino acids.

STIMULATE or STIMULATING, in relationship to the term "response" shall mean that a response is increased in the presence of a compound as opposed to in the
5 absence of the compound.

SUBSTANTIALLY shall refer to a result which is within 40% of a control result, preferably within 35%, more preferably within 30%, more preferably within 25%, more preferably within 20%, more preferably within 15%, more preferably within 10%, more preferably within 5%, more preferably within 2%, and most preferably within 1% of a
10 control result. For example, in the context of receptor functionality, a test receptor may exhibit substantially similar results to a control receptor if the transduced signal, measured using a method taught herein or similar method known to the art-skilled, is within 40% of the signal produced by a control signal.

VECTOR in reference to cDNA shall mean a circular DNA capable of
15 incorporating at least one cDNA and capable of incorporation into a Host Cell.

The order of the following sections is set forth for presentational efficiency and is not intended, nor should be construed, as a limitation on the disclosure or the claims to follow.

20 A. Introduction

The traditional study of receptors has typically proceeded from the *a priori* assumption (historically based) that the endogenous ligand must first be identified before discovery could proceed to find antagonists and other molecules that could affect the receptor. Even in cases where an antagonist might have been known first, the search
25 immediately extended to looking for the endogenous ligand. This mode of thinking has persisted in receptor research even after the discovery of constitutively activated receptors. What has not been heretofore recognized is that it is the active state of the receptor that is most useful for discovering agonists and inverse agonists of the receptor. For those diseases which result from an overly active receptor or an under-active receptor, what is desired in a
30 therapeutic drug is a compound which acts to diminish the active state of a receptor or enhance the activity of the receptor, respectively, not necessarily a drug which is an antagonist to the endogenous ligand. This is because a compound that reduces or enhances

AREN-0309**PATENT**

the activity of the active receptor state need not bind at the same site as the endogenous ligand. Thus, as taught by a method of this invention, any search for therapeutic compounds should start by screening compounds against the ligand-independent active state.

5 **B. Identification of Human GPCRs**

The efforts of the Human Genome project has led to the identification of a plethora of information regarding nucleic acid sequences located within the human genome; it has been the case in this endeavor that genetic sequence information has been made available without an understanding or recognition as to whether or not any particular genomic
 10 sequence does or may contain open-reading frame information that translate human proteins. Several methods of identifying nucleic acid sequences within the human genome are within the purview of those having ordinary skill in the art. For example, and not limitation, a variety of human GPCRs, disclosed herein, were discovered by reviewing the GenBank™ database. Table B, below, lists several endogenous GPCRs that we have
 15 discovered, along with other GPCRs that are homologous to the disclosed GPCR.

TABLE B

Disclosed Human Orphan GPCRs	Accession Number Identified	Open Reading Frame (Base Pairs)	Reference To Homologous GPCR	Per Cent Homology To Designated GPCR
hRUP28	AC073957	1,002bp	hGPR30	34%
hRUP29	AC083865	918bp	hGPR18	27%
hRUP30	AC055863	1,125bp	hBRB1	27%
hRUP31	AL356214	1,086bp	hGALR-1	31%
hRUP32	AL513524	1,038bp	hPNR	43%
hRUP33	AL513524	1,020bp	GPR57 GPR58	50% 51%
hRUP34	AL513524	1,029bp	hPNR	45%
hRUP35	AC021089	1,062bp	hk-type 3 opioid	27%
hRUP36	AC090099	969bp	GPR90	42%

hRUP37	AC090099	969bp	hMRG	41%
--------	----------	-------	------	-----

Receptor homology is useful in terms of gaining an appreciation of a role of the receptors within the human body. As the patent document progresses, techniques for mutating these receptors to establish non-endogenous, constitutively activated versions of these receptors will be discussed.

The techniques disclosed herein are also applicable to other human GPCRs known to the art, as will be apparent to those skilled in the art.

C. Receptor Screening

Screening candidate compounds against a non-endogenous, constitutively activated version of the GPCRs disclosed herein allows for the direct identification of candidate compounds which act at the cell surface receptor, without requiring use of the receptor's endogenous ligand. Using routine, and often commercially available techniques, one can determine areas within the body where the endogenous version of human GPCRs disclosed herein is expressed and/or over-expressed. The expression location of a receptor in a specific tissue provides a scientist with the ability to assign a physiological functional role of the receptor. It is also possible using these techniques to determine related disease/disorder states which are associated with the expression and/or over-expression of the receptor; such an approach is disclosed in this patent document. Furthermore, expression of a receptor in diseased organs can assist one in determining the magnitude of the clinical relevance of the receptor.

Constitutive activation of the GPCRs disclosed herein is based upon the distance from the proline residue at which is presumed to be located within TM6 of the GPCR; this algorithmic technique is disclosed in co-pending and commonly assigned patent document PCT Application Number PCT/US99/23938, published as WO 00/22129 on April 20, 2000, which, along with the other patent documents listed herein, is incorporated herein by reference. The algorithmic technique is not predicated upon traditional sequence "alignment" but rather a specified distance from the aforementioned TM6 proline residue (or, of course, endogenous constitutive substitution for such proline residue). By mutating the amino acid residue located 16 amino acid residues from this residue (presumably located in the IC3 region of the receptor) to, most preferably, a lysine residue, constitutive

activation of the receptor may be obtained. Other amino acid residues may be useful in the mutation at this position to achieve this objective and will be discussed in detail, below.

D. Disease/Disorder Identification and/or Selection

As will be set forth in greater detail below, inverse agonists and agonists to the non-
5 endogenous, constitutively activated GPCR can be identified by the methodologies of this invention. Such inverse agonists and agonists are ideal candidates as lead compounds in drug discovery programs for treating diseases related to this receptor. Because of the ability to directly identify inverse agonists to the GPCR, thereby allowing for the development of pharmaceutical compositions, a search for diseases and disorders associated with the GPCR
10 is relevant. The expression location of a receptor in a specific tissue provides a scientist with the ability to assign a physiological function to the receptor. For example, scanning both diseased and normal tissue samples for the presence of the GPCR now becomes more than an academic exercise or one which might be pursued along the path of identifying an endogenous ligand to the specific GPCR. Tissue scans can be conducted across a broad
15 range of healthy and diseased tissues. Such tissue scans provide a potential first step in associating a specific receptor with a disease and/or disorder. Furthermore, expression of a receptor in diseased organs can assist one in determining the magnitude of clinical relevance of the receptor.

The DNA sequence of the GPCR can be used to make a probe/primer. In some
20 preferred embodiments the DNA sequence is used to make a probe for (a) dot-blot analysis against tissue-mRNA, and/or (b) RT-PCR identification of the expression of the receptor in tissue samples. The presence of a receptor in a tissue source, or a diseased tissue, or the presence of the receptor at elevated concentrations in diseased tissue compared to a normal tissue, can be used to correlate location to function and indicate the receptor's physiological
25 role/function and create a treatment regimen, including but not limited to, a disease associated with that function/role. Receptors can also be localized to regions of organs by this technique. Based on the known or assumed roles/functions of the specific tissues to which the receptor is localized, the putative physiological function of the receptor can be deduced. For example and not limitation, proteins located/expressed in areas of the
30 thalamus are associated with sensorimotor processing and arousal (*see*, Goodman & Gilman's, The Pharmacological Basis of Therapeutics, 9th Edition, page 465 (1996)). Proteins expressed in the hippocampus or in Schwann cells are associated with learning and

memory, and myelination of peripheral nerves, respectively (*see*, Kandel, E. et al., Essentials of Neural Science and Behavior pages 657, 680 and 28, respectively (1995)).

E. Screening of Candidate Compounds

5 1. Generic GPCR screening assay techniques

When a G protein receptor becomes constitutively active, it binds to a G protein (*e.g.*, G_q, G_s, G_i, G_z, G_o) and stimulates the binding of GTP to the G protein. The G protein then acts as a GTPase and hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes deactivated. However, constitutively activated receptors continue to
10 exchange GDP to GTP. A non-hydrolyzable analog of GTP, [³⁵S]GTPγS, can be used to monitor enhanced binding to membranes which express constitutively activated receptors. It is reported that [³⁵S]GTPγS can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and available to those in the art, was reported by Traynor and Nahorski in
15 1995. The use of this assay system is typically for initial screening of candidate compounds because the system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

2. Specific GPCR screening assay techniques

Once candidate compounds are identified using the “generic” G protein-coupled
20 receptor assay (*i.e.*, an assay to select compounds that are agonists or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the “generic” assay may not bind to the receptor, but may instead merely “uncouple” the G protein from the intracellular domain.

a. G_s, G_z and G_i.

25 G_s stimulates the enzyme adenylyl cyclase. G_i (and G_z and G_o), on the other hand, inhibits adenylyl cyclase. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus, constitutively activated GPCRs that couple the G_s protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple G_i (or G_z, G_o) protein are associated with decreased cellular levels of cAMP. *See*,
30 generally, “Indirect Mechanisms of Synaptic Transmission,” Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can be utilized to determine if a candidate compound is, *e.g.*, an inverse

agonist to the receptor (*i.e.*, such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be utilized is a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) that then binds to the promoter at specific sites (cAMP response elements) and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, *e.g.*, β -galactosidase or luciferase. Thus, a constitutively activated G_s -linked receptor causes the accumulation of cAMP that then activates the gene and leads to the expression of the reporter protein. The reporter protein such as β -galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995).

15 **b. G_o and G_q**

G_q and G_o are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP_2 , releasing two intracellular messengers: diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3). Increased accumulation of IP_3 is associated with activation of G_q - and G_o -associated receptors. *See, generally*, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Assays that detect IP_3 accumulation can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to a G_q - or G_o -associated receptor (*i.e.*, such a compound would decrease the levels of IP_3). G_q -associated receptors can also be examined using an AP1 reporter assay wherein G_q -dependent phospholipase C causes activation of genes containing AP1 elements; thus, activated G_q -associated receptors will evidence an increase in the expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will evidence an increase in such expression. Commercially available assays for such detection are available.

30 **3. GPCR Fusion Protein**

The use of an endogenous, constitutively activated GPCR or a non-endogenous, constitutively activated GPCR, for use in screening of candidate compounds for the direct identification of inverse agonists, agonists provide an interesting screening challenge in that, by definition, the receptor is active even in the absence of an endogenous ligand bound thereto. Thus, in order to differentiate between, *e.g.*, the non-endogenous receptor in the presence of a candidate compound and the non-endogenous receptor in the absence of that compound, with an aim of such a differentiation to allow for an understanding as to whether such compound may be an inverse agonist or agonist or have no affect on such a receptor, it is preferred that an approach be utilized that can enhance such differentiation. A preferred approach is the use of a GPCR Fusion Protein.

Generally, once it is determined that a non-endogenous GPCR has been constitutively activated using the assay techniques set forth above (as well as others), it is possible to determine the predominant G protein that couples with the endogenous GPCR. Coupling of the G protein to the GPCR provides a signaling pathway that can be assessed. In some embodiments it is preferred that screening take place using a mammalian expression system, such a system will be expected to have endogenous G protein therein. Thus, by definition, in such a system, the non-endogenous, constitutively activated GPCR will continuously signal. In some embodiments it is preferred that this signal be enhanced such that in the presence of, *e.g.*, an inverse agonist to the receptor, it is more likely that it will be able to more readily differentiate, particularly in the context of screening, between the receptor when it is contacted with the inverse agonist.

The GPCR Fusion Protein is intended to enhance the efficacy of G protein coupling with the non-endogenous GPCR. The GPCR Fusion Protein is preferred for screening with either an endogenous, constitutively active GPCR or a non-endogenous, constitutively activated GPCR because such an approach increases the signal that is utilized in such screening techniques. This is important in facilitating a significant "signal to noise" ratio; such a significant ratio is preferred for the screening of candidate compounds as disclosed herein.

The construction of a construct useful for expression of a GPCR Fusion Protein is within the purview of those having ordinary skill in the art. Commercially available expression vectors and systems offer a variety of approaches that can fit the particular needs of an investigator. Important criteria on the construction of such a GPCR Fusion Protein

construct include but are not limited to, that the endogenous GPCR sequence and the G protein sequence both be in-frame (preferably, the sequence for the endogenous GPCR is upstream of the G protein sequence), and that the "stop" codon of the GPCR be deleted or replaced such that upon expression of the GPCR, the G protein can also be expressed.

- 5 Other embodiments include constructs wherein the endogenous GPCR sequence and the G protein sequence are not in-frame and/or the "stop" codon is not deleted or replaced. The GPCR can be linked directly to the G protein, or there can be spacer residues between the two (preferably, no more than about 12, although this number can be readily ascertained by one of ordinary skill in the art). Based upon convenience it is preferred to use a spacer.
- 10 Preferably, the G protein that couples to the non-endogenous GPCR will have been identified prior to the creation of the GPCR Fusion Protein construct. Because there are only a few G proteins that have been identified, it is preferred that a construct comprising the sequence of the G protein (*i.e.*, a universal G protein construct (*see Examples*)) be available for insertion of an endogenous GPCR sequence therein; this provides for further
- 15 efficiency in the context of large-scale screening of a variety of different endogenous GPCRs having different sequences.

As noted above, constitutively activated GPCRs that couple to G_i , G_z and G_o are expected to inhibit the formation of cAMP making assays based upon these types of GPCRs challenging (*i.e.*, the cAMP signal decreases upon activation thus making the direct

20 identification of, *e.g.*, inverse agonists (which would further decrease this signal), challenging. As will be disclosed herein, we have ascertained that for these types of receptors, it is possible to create a GPCR Fusion Protein that is not based upon the GPCRs endogenous G protein, in an effort to establish a viable cyclase-based assay. Thus, for example, an endogenous G_i coupled receptor can be fused to a G_s protein—such a fusion

25 construct, upon expression, "drives" or "forces" the endogenous GPCR to couple with, *e.g.*, G_s rather than the "natural" G_i protein, such that a cyclase-based assay can be established. Thus, for G_i , G_z and G_o coupled receptors, in some embodiments it is preferred that when a GPCR Fusion Protein is used and the assay is based upon detection of adenylyl cyclase activity, that the fusion construct be established with G_s (or an equivalent G protein that

30 stimulates the formation of the enzyme adenylyl cyclase).

G	Effect of cAMP	Effect of IP_3	Effect of	Effect on IP_3
---	----------------	------------------	-----------	------------------

protein	Production upon Activation of GPCR (<i>i.e.</i> , constitutive activation or agonist binding)	Accumulation upon Activation of GPCR (<i>i.e.</i> , constitutive activation or agonist binding)	cAMP Production upon contact with an Inverse Agonist	Accumulation upon contact with an Inverse Agonist
G _s	Increase	N/A	Decrease	N/A
G _i	Decrease	N/A	Increase	N/A
G _z	Decrease	N/A	Increase	N/A
G _o	Decrease	Increase	Increase	Decrease
G _q	N/A	Increase	N/A	Decrease

Equally effective is a G Protein Fusion construct that utilizes a G_q Protein fused with a G_s, G_i, G_z or G_o Protein. In some embodiments a preferred fusion construct can be accomplished with a G_q Protein wherein the first six (6) amino acids of the G-protein α -subunit ("G α q") is deleted and the last five (5) amino acids at the C-terminal end of G α q is replaced with the corresponding amino acids of the G α of the G protein of interest. For example, a fusion construct can have a G_q (6 amino acid deletion) fused with a G_i Protein, resulting in a "G_q/G_i Fusion Construct". This fusion construct will forces the endogenous G_i coupled receptor to couple to its non-endogenous G protein, G_q, such that the second messenger, for example, inositol triphosphate or diacylglycerol, can be measured *in lieu* of cAMP production.

4. Co-transfection of a Target G_i Coupled GPCR with a Signal-Enhancer G_s Coupled GPCR (cAMP Based Assays)

A G_i coupled receptor is known to inhibit adenylyl cyclase, and, therefore, decreases the level of cAMP production, which can make assessment of cAMP levels challenging. An effective technique in measuring the decrease in production of cAMP as an indication of constitutive activation of a receptor that predominantly couples G_i upon activation can be accomplished by co-transfecting a signal enhancer, *e.g.*, a non-endogenous, constitutively activated receptor that predominantly couples with G_s upon activation (*e.g.*, TSHR-A623I, disclosed below), with the G_i linked GPCR. As is apparent, constitutive activation of a G_s

coupled receptor can be determined based upon an increase in production of cAMP. Constitutive activation of a G_i coupled receptor leads to a decrease in production cAMP. Thus, the co-transfection approach is intended to advantageously exploit these "opposite" affects. For example, co-transfection of a non-endogenous, constitutively activated G_s coupled receptor (the "signal enhancer") with the endogenous G_i coupled receptor (the "target receptor") provides a baseline cAMP signal (*i.e.*, although the G_i coupled receptor will decrease cAMP levels, this "decrease" will be relative to the substantial increase in cAMP levels established by constitutively activated G_s coupled signal enhancer). By then co-transfecting the signal enhancer with a constitutively activated version of the target receptor, cAMP would be expected to further decrease (relative to base line) due to the increased functional activity of the G_i target (*i.e.*, which decreases cAMP).

Screening of candidate compounds using a cAMP based assay can then be accomplished, with two 'changes' relative to the use of the endogenous receptor/G-protein fusion: first, relative to the G_i coupled target receptor, "opposite" effects will result, *i.e.*, an inverse agonist of the G_i coupled target receptor will increase the measured cAMP signal, while an agonist of the G_i coupled target receptor will decrease this signal; second, as would be apparent, candidate compounds that are directly identified using this approach should be assessed independently to ensure that these do not target the signal enhancing receptor (this can be done prior to or after screening against the co-transfected receptors).

20

F. Medicinal Chemistry

Generally, but not always, direct identification of candidate compounds is conducted in conjunction with compounds generated via combinatorial chemistry techniques, whereby thousands of compounds are randomly prepared for such analysis. Generally, the results of such screening will be compounds having unique core structures; thereafter, these compounds may be subjected to additional chemical modification around a preferred core structure(s) to further enhance the medicinal properties thereof. Such techniques are known to those in the art and will not be addressed in detail in this patent document.

30

G. Pharmaceutical compositions

Candidate compounds selected for further development can be formulated into pharmaceutical compositions using techniques well known to those in the art. Suitable pharmaceutically-acceptable carriers are available to those in the art; for example, see Remington's Pharmaceutical Sciences, 16th Edition, 1980, Mack Publishing Co., (Osol et al., eds.).

H. Other Utilities

Although a preferred use of the non-endogenous versions of the GPCRs disclosed herein may be for the direct identification of candidate compounds as inverse agonists or agonists (preferably for use as pharmaceutical agents), other uses of these versions of GPCRs exist. For example, *in vitro* and *in vivo* systems incorporating GPCRs can be utilized to further elucidate and understand the roles these receptors play in the human condition, both normal and diseased, as well as understanding the role of constitutive activation as it applies to understanding the signaling cascade. In some embodiments it is preferred that the endogenous receptors be "orphan receptors", *i.e.*, the endogenous ligand for the receptor has not been identified. In some embodiments, therefore, the modified, non-endogenous GPCRs can be used to understand the role of endogenous receptors in the human body before the endogenous ligand therefore is identified. Such receptors can also be used to further elucidate known receptors and the pathways through which they transduce a signal. Other uses of the disclosed receptors will become apparent to those in the art based upon, *inter alia*, a review of this patent document.

EXAMPLES

The following examples are presented for purposes of elucidation, and not limitation, of the present invention. While specific nucleic acid and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor modifications to these sequences while achieving the same or substantially similar results reported below. The traditional approach to application or understanding of sequence cassettes from one sequence to another (*e.g.* from rat receptor to human receptor or from human receptor A to human receptor B) is generally predicated upon sequence alignment techniques whereby the sequences are aligned in an effort to determine areas of commonality. The mutational approach disclosed herein does not rely upon this approach but is instead based upon an algorithmic approach and a positional distance from a conserved proline residue located within the TM6 region of human GPCRs. Once this

AREN-0309**PATENT**

approach is secured, those in the art are credited with the ability to make minor modifications thereto to achieve substantially the same results (*i.e.*, constitutive activation) disclosed herein. Such modified approaches are considered within the purview of this disclosure.

5 Example 1**ENDOGENOUS HUMAN GPCRS****1. Identification of Human GPCRs**

The disclosed endogenous human GPCRs were identified based upon a review of the GenBank™ database information. While searching the database, the following
10 cDNA clones were identified as evidenced below (Table C).

TABLE C

Disclosed Human Orphan GPCRs	Accession Number Identified	Open Reading Frame (Base Pairs)	Reference To Homologous GPCR	Nucleic Acid SEQ.ID. NO.	Amino Acid SEQ.ID.NO.
hRUP28	AC073957	1,002bp	hGPR30	1	2
hRUP29	AC083865	918bp	hGPR18	3	4
hRUP30	AC055863	1,125bp	hBRB1	5	6
hRUP31	AL356214	1,086bp	hGALR-1	7	8
hRUP32	AL513524	1,038bp	hPNR	9	10
hRUP33	AL513524	1,020bp	GPR57 GPR58	11	12
hRUP34	AL513524	1,029bp	hPNR	13	14
hRUP35	AC021089	1,062bp	hκ-type 3 opioid	15	16
hRUP36	AC090099	969bp	GPR90	17	18
hRUP37	AC090099	969bp	hMRG	19	20

2. Full Length Cloning**a. hRUP28 (Seq. Id. Nos. 1 & 2)**

15 The disclosed human RUP28 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession Number AC073957 was identified as a human genomic sequence from chromosome 7.

The full length RUP28 was cloned by PCR using primers:

5'-

CAGAGCTCTGGTGGCCACCTCTGTCC-

3' (SEQ.ID.NO.:21; sense, 5' of initiation codon),

5'-CTGCGTCCACCAGAGTCACGTCTCC-3' (SEQ.ID.NO.:22; antisense, 3' of stop codon), and human adult liver Marathon-Ready™ cDNA (Clontech) as template.

- 5 Advantage™ cDNA polymerase (Clontech) was used for the amplification in a 50µl reaction by the following cycle with step 2 to 4 repeated 35 times: 95°C for 5 min; 94°C for 30 sec; 58°C for 30 sec; 72°C for 1 min 30 sec; and 72°C for 7 min.

A 1.16kb PCR fragment was isolated from a 1% agarose gel and cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator Kit (P.E. Biosystems). See, SEQ.ID.NO.:1 for the nucleic acid sequence and SEQ.ID.NO.:2 for the putative amino acid sequence.

b. hRUP29 (Seq. Id. Nos. 3 & 4)

The disclosed human RUP29 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession Number AC0083865 was identified as a human genomic sequence from chromosome 7.

The full length RUP29 was cloned by PCR using primers:

5'-GTATGCCTGGCCACAATACCTCCAGG-3' (SEQ.ID.NO.:23; sense, containing the initiation codon),

- 5'-GTTTGTGGCTAACGGCACAAAACACAATTCC-3' (SEQ.ID.NO.:24; antisense, containing the stop codon) and human genomic DNA as template. TaqPlus® Precision DNA polymerase (Stratagene) was used for the amplification in a 50µl reaction by the following cycle with step 2 to 4 repeated 35 times: 94°C for 5 min; 94°C for 30 sec; 54°C for 30 sec; 72°C for 1 min 30 sec; and 72°C for 7 min.

A 930bp PCR fragment was isolated from a 1% agarose gel and cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator Kit (P.E. Biosystems).

Rapid amplification of cDNA ends (RACE) was performed using human leukocyte and ovary Marathon-Ready™ cDNA (Clontech) to determine the precise 5' end of RUP29 cDNA. RUP29 specific primer (1) having the sequence:

- 30 5-GGTACCACAATGACAATCACCAGCGTCC-3' (SEQ.ID.NO.:25)

and AP1 primer (Clontech) were used for the first-round PCR reaction, and RUP29 specific primer (2) having the following sequence:

5'-

GGAACGTGAGGTACATGTGGATGTGCAGC-3' (SEQ.ID.NO.:26)

and AP2 primer (Clontech) were used for the second-round PCR reaction. The products of the RACE reactions were isolated and cloned into the pCRII-TOPO vector
5 (Invitrogen) and sequenced. See, SEQ.ID.NO.:3 for the nucleic acid sequence and SEQ.ID.NO.:4 for the putative amino acid sequence.

c. hRUP30 (Seq. Id. Nos. 5 & 6)

The disclosed human RUP30 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession
10 Number AC055863 was identified as a human genomic sequence from chromosome 17.

The full length RUP30 was cloned by 5'RACE -PCR with a human pancreas Marathon-Ready™ cDNA (Clontech) as template and the following oligonucleotide:

5'-GCAGTGTAGCGGTCAACCGTGAGCAGG-3'(SEQ.ID.NO.:27; sense, containing the initiation codon), and AP1 primer (Clontech) were used for the first round of RT-PCR and
15 oligonucleotide:

5'-TGAGCAGGATGGCGATCCAGACTGAGGCGTGG-3'(SEQ.ID.NO.:28; antisense, containing the stop codon) and AP2 primer (Clontech) were used for the second round of PCR. DNA fragments generated by the 5' RACE-PCR were cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the SP6/T7 primers (Stratagene).

20 Based on the sequence of the 5' RACE products, the full length RUP30 was cloned by RT-PCR, using primers:

5'-GAGGTACAGCTGGCGATGCTGACAG-3' (SEQ.ID.NO.:29; sense, ATG as the initiation codon);

5'-GTGGCCATGAGCCACCCTGAGCTCC-3' (SEQ.ID.NO.:30; antisense, 3' of the stop
25 codon) and human pancreas Marathon-Ready™ cDNA (Clontech) as template. Taq DNA polymerase (Stratagene) was used for the amplification in 50 µl reaction by the following cycle with step 2 to step 4 repeated 35 times: 94°C for 40 seconds; 94°C for 20 seconds; 64°C for 20 seconds; 72°C for 2 minutes; and 72°C for 5 minutes.

A 1.2 Kb PCR fragment was isolated from a 1% agarose gel and cloned into the
30 pCRII-TOPO vector (Invitrogen) and several clones were sequenced using the ABI Big Dye Terminator kit (P.E. Biosystems). See, SEQ.ID.NO.:5 for the nucleic acid sequence and SEQ.ID.NO.:6 for the putative amino acid sequence.

d. hRUP31 (Seq. Id. Nos. 7 & 8)

The disclosed human RUP31 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession Number AL356214 was identified as a human genomic sequence from chromosome 10.

5 The full length RUP31 was cloned by RT-PCR using primers:

5'-GGAATGTCCACTGAATGCGCGCGG-3' (SEQ.ID.NO.:31; sense, containing the initiation codon),

5'-AGCTCGCCAGGTGTGAGAACTCGG-3' (SEQ.ID.NO.:32; antisense, 3' of stop codon) and human mammary gland Marathon-Ready™ cDNA (Clontech) as template.

10 Advantage™ cDNA polymerase (Clontech) was used for the amplification in 50 µl reaction by the following cycle with step 2 to step 4 repeated 35 times: 94°C for 40 sec; 94°C for 20 sec; 66°C for 20 sec; 72°C for 1 min 30 sec; and 72°C for 5 min.

A 1.1 kb PCR fragment was isolated from a 1% agarose gel and cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator Kit
15 (P.E. Biosystems). *See*, SEQ.ID.NO.:7 for the nucleic acid sequence and SEQ.ID.NO.:8 for the putative amino acid sequence.

e. hRUP32 (Seq. Id. Nos. 9 & 10)

The disclosed human RUP32 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession
20 Number AL513524 was identified as a human genomic sequence from chromosome 6.

The full length RUP32 was cloned by PCR using primers:

5'-GCGTTATGAGCAGCAATTCATCCCTGCTGG-3' (SEQ.ID.NO.:33; sense, containing the initiation codon),

5'-GTATCCTGAACTTCGTCTATACTGTC-3' (SEQ.ID.NO.:34; antisense)

25 and human genomic DNA (Clontech) as template. TaqPlus® Precision DNA polymerase (Stratagene) was used for the amplification by the following cycle with step 2 to step 4 repeated 35 times: 94°C for 3 min; 94°C for 20 sec; 58°C for 20 sec; 72°C for 1 min 30 sec; and 72°C for 7 min.

A 1.06 kb PCR fragment was isolated from a 1% agarose gel and cloned into the
30 pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator Kit (P.E. Biosystems). *See*, SEQ.ID.NO.:9 for the nucleic acid sequence and SEQ.ID.NO.:10 for the putative amino acid sequence.

f. hRUP33 (Seq. Id. Nos. 11 & 12)

The disclosed human RUP33 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession Number AL513524 was identified as a human genomic sequence from chromosome 6.

5 The full length RUP33 was cloned by PCR using primers:

5'-CCCTCAGGAATGATGCCCTTTGCCACAA-3' (SEQ.ID.NO.:35; sense, containing the initiation codon),

5'-ATCCATGTGGTTGGTGCATGTGGTTCGT-3' (SEQ.ID.NO.:36; antisense)

and human genomic DNA (Clontech) as template. TaqPlus® Precision DNA polymerase
10 (Stratagene) was used for the amplification by the following cycle with step 2 to step 4 repeated 35 times: 94°C for 3 min; 94°C for 20 sec; 56°C for 20 sec; 72°C for 1 min 30 sec; and 72°C for 7 min.

A 1.1 kb PCR fragment was isolated from a 1% agarose gel and cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator Kit
15 (P.E. Biosystems). See, SEQ.ID.NO.:11 for the nucleic acid sequence and SEQ.ID.NO.:12 for the putative amino acid sequence.

g. hRUP34 (Seq. Id. Nos. 13 & 14)

The disclosed human RUP34 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession
20 Number AL513524 was identified as a human genomic sequence from chromosome 6.

The full length RUP34 was cloned by PCR using primers:

5'-AAACAACAAACAGCAGAACCATGACCAGC-3' (SEQ.ID.NO.:37; sense, containing the initiation codon),

5'-ACATAGAGACAAGTGACATGTGTGAACCAC-3' (SEQ.ID.NO.:38; antisense)

25 and human genomic DNA (Clontech) as template. TaqPlus® Precision DNA polymerase (Stratagene) was used for the amplification by the following cycle with step 2 to step 4 repeated 35 times: 94°C for 3 min; 94°C for 20 sec; 60°C for 20 sec; 72°C for 1 min 30 sec; and 72°C for 7 min.

A 1.27 kb PCR fragment was isolated from a 1% agarose gel and cloned into the
30 pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator Kit (P.E. Biosystems). See, SEQ.ID.NO.:13 for the nucleic acid sequence and SEQ.ID.NO.:14 for the putative amino acid sequence.

h. hRUP35 (Seq. Id. Nos. 15 & 16)

The disclosed human RUP35 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession Number AC021089 was identified as a human genomic sequence from chromosome 16.

5 The 5' sequence of RUP35 was determined by 5' RACE-PCR with a human fetal brain Marathon-Ready™ cDNA (Clontech) as template. Oligonucleotide 5'-GGTATGAGACCGTGTGGTACTTGAGC-3' (SEQ.ID.NO.:39; sense) and AP1 primer (Clontech) were used for the first round of RT-PCR and oligonucleotide 5'-GTGGCAGACAGCGATATACCTGTCAATGG-3' (SEQ.ID.NO.:40; antisense) 10 and AP2 primer (Clontech) were used for the second round of PCR. DNA fragments generated by the 5' RACE-PCR were cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the SP6/T7 primers (Stratagene).

Based upon the sequence of the 5' RACE products, the full length RUP35 was cloned by RT-PCR, using primers

15 5'-GCGCTCATGGAGCACACGCACGCCAC-3' (SEQ.ID.NO.:41; sense, ATG as the initiation codon) and 5'-GAGGCAGTAGTTGCCACACCTATGG-3' (SEQ.ID.NO.:42; antisense, 3' of the stop codon) and human brain Marathon-Ready™ cDNA (Clontech) as template. Advantage™ cDNA polymerase (Clontech) was used for the amplification in 100 µl reaction by the 20 following cycle with step 2 to step 4 repeated 45 times: 95°C for 2 min; 95°C for 20 sec; 60°C for 20 sec; 72°C for 1 min 30 sec; and 72°C for 5 min.

A 1.0 kb PCR fragment was isolated from a 1% agarose gel and cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator Kit (P.E. Biosystems). See, SEQ.ID.NO.:15 for the nucleic acid sequence and 25 SEQ.ID.NO.:16 for the putative amino acid sequence.

i. hRUP36 (Seq. Id. Nos. 17 & 18)

The disclosed human RUP36 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession Number AC090099 was identified as a human genomic sequence from chromosome 11.

30 The full length RUP36 was cloned by PCR using primers: 5'-CATCTGGTTTGTGTTCCAGGGGCACCAG-3' (SEQ.ID.NO.:43; sense, 5' of start codon),

5'-

GACAGTGTGCTCTCAAAGTCCCGTCTGACTG -3' (SEQ.ID.NO.:44; antisense, 3' of stop codon) and human genomic DNA (Clontech) as template. TaqPlus® Precision DNA polymerase (Stratagene) was used for the amplification in a 50µl reaction by the following cycle with step 2 to step 4 repeated 30 times: 95°C, 5 min; 95°C for 30 sec; 70°C for 30 sec; 72°C for 1 min 30 sec; and 72°C for 7 min.

A 1.0 kb PCR fragment was isolated from a 1% agarose gel and cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator Kit (P.E. Biosystems). See, SEQ.ID.NO.:17 for the nucleic acid sequence and SEQ.ID.NO.:18 for the putative amino acid sequence.

j. hRUP37 (Seq. Id. Nos. 19 & 20)

The disclosed human RUP37 was identified based upon the use of GenBank database information. While searching the database, a cDNA clone with Accession Number AC090099 was identified as a human genomic sequence from chromosome 11.

The full length RUP37 was cloned by PCR using primers:

5'-CTGTTTCCAGGGTCATCAGACTGGG-3' (SEQ.ID.NO.:45; sense);

5'-GCAGCATTGCTCTCAAAGTCCTGTCTG-3' (SEQ.ID.NO.:46; antisense)

and human genomic DNA (Clontech) as template. TaqPlus® Precision DNA polymerase (Stratagene) was used for the amplification by the following cycle with step 2 to step 4 repeated 35 times: 95°C for 5 min; 95°C for 30 sec; 62°C for 30 sec; 72°C for 1 min 30 sec; and 72°C for 7 min.

A 969 base pair was isolated from a 1%% agarose gel and cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator Kit (P.E. Biosystems). See, SEQ.ID.NO.:19 for the nucleic acid sequence and SEQ.ID.NO.:20 for the putative amino acid sequence.

Example 2

PREPARATION OF NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED GPCRS

Those skilled in the art are credited with the ability to select techniques for mutation of a nucleic acid sequence. Presented below are approaches utilized to create non-endogenous versions of several of the human GPCRs disclosed above. The mutations disclosed below are based upon an algorithmic approach whereby the 16th amino acid (located in the IC3 region of the GPCR) from a conserved proline (or an

endogenous, conservative substitution therefore) residue (located in the TM6 region of the GPCR, near the TM6/IC3 interface) is mutated, preferably to an alanine, histidine, arginine or lysine amino acid residue, most preferably to a lysine amino acid residue.

5 1. Transformer Site-Directed™ Mutagenesis

Preparation of non-endogenous human GPCRs may be accomplished on human GPCRs using, *inter alia*, Transformer Site-Directed™ Mutagenesis Kit (Clontech) according to the manufacturer instructions. In some embodiments two mutagenesis primers are used, preferably a lysine mutagenesis oligonucleotide that creates the lysine mutation, and a selection marker oligonucleotide. For convenience, the codon mutation to be incorporated into the human GPCR is also noted, in standard form (Table D):

TABLE D

Receptor Identifier	Codon Mutation
hRUP28	V274K
hRUP29	T249K
hRUP30	R232K
hRUP31	M294K
hRUP32	F220K
hRUP34	A238K
hRUP35	Y215K
hRUP36	L294K
hRUP37	T219K

15 Example 3

RECEPTOR EXPRESSION

Although a variety of cells are available to the art-skilled for the expression of proteins, it is preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, *i.e.*, utilization of, *e.g.*, yeast cells for the expression of a GPCR, while possible, introduces into the protocol a non-mammalian cell which may not (indeed, in the case of yeast, does not) include the receptor-coupling, genetic-mechanism

and secretory pathways that have evolved for mammalian systems – thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as those obtained using mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular needs of the artisan.

a. Transient Transfection

On day one, 6×10^6 cells/10 cm dish of 293 cells well were plated out. On day two, two reaction tubes were prepared (the proportions to follow for each tube are per plate): tube A was prepared by mixing 4 μ g DNA (*e.g.*, pCMV vector; pCMV vector with receptor cDNA, *etc.*) in 0.5 ml serum free DMEM (Gibco BRL); tube B was prepared by mixing 24 μ l lipofectamine (Gibco BRL) in 0.5ml serum free DMEM. Tubes A and B were admixed by inversion (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the “transfection mixture”. Plated 293 cells were washed with 1XPBS, followed by addition of 5 ml serum free DMEM. One ml of the transfection mixture were added to the cells, followed by incubation for 4hrs at 37°C/5% CO₂. The transfection mixture was removed by aspiration, followed by the addition of 10ml of DMEM/10% Fetal Bovine Serum. Cells were incubated at 37°C/5% CO₂. After 48hr incubation, cells were harvested and utilized for analysis.

b. Stable Cell Lines

Approximately 12×10^6 293 cells will be plated on a 15cm tissue culture plate, and grown in DME High Glucose Medium containing 10% fetal bovine serum and one percent sodium pyruvate, L-glutamine, and antibiotics. Twenty-four hours following plating of 293 cells (to approximately ~80% confluency), the cells will be transfected using 12 μ g of DNA. The 12 μ g of DNA is combined with 60 μ l of lipofectamine and 2mL of DME High Glucose Medium without serum. The medium will be aspirated from the plates and the cells washed once with medium without serum. The DNA, lipofectamine, and medium mixture will be added to the plate along with 10mL of medium without serum. Following incubation at 37°C for four to five hours, the medium will be aspirated and 25ml of medium containing serum will be added. Twenty-four hours following transfection, the medium will be aspirated again, and fresh medium with serum will be added. Forty-eight hours following transfection, the medium will be aspirated and medium with serum will be added containing geneticin (G418 drug) at a final concentration of 500 μ g/mL. The transfected cells will then

undergo selection for positively transfected cells containing the G418 resistant gene. The medium will be replaced every four to five days as selection occurs. During selection, cells will be grown to create stable pools, or split for stable clonal selection.

Example 4

5 ASSAYS FOR DETERMINATION OF CONSTITUTIVE ACTIVITY OF NON-ENDOGENOUS GPCRS

A variety of approaches are available for assessment of constitutive activity of the non-endogenous human GPCRs. The following are illustrative; those of ordinary skill in
10 the art are credited with the ability to determine those techniques that are preferentially beneficial for the needs of the artisan.

1. Membrane Binding Assays: [³⁵S]GTPγS Assay

When a G protein-coupled receptor is in its active state, either as a result of ligand
15 binding or constitutive activation, the receptor couples to a G protein and stimulates the release of GDP and subsequent binding of GTP to the G protein. The alpha subunit of the G protein-receptor complex acts as a GTPase and slowly hydrolyzes the GTP to GDP, at which point the receptor normally is deactivated. Constitutively activated receptors continue to exchange GDP for GTP. The non-hydrolyzable GTP analog, [³⁵S]GTPγS, can
20 be utilized to demonstrate enhanced binding of [³⁵S]GTPγS to membranes expressing constitutively activated receptors. Advantages of using [³⁵S]GTPγS binding to measure constitutive activation include but are not limited to the following: (a) it is generically applicable to all G protein-coupled receptors; (b) it is proximal at the membrane surface making it less likely to pick-up molecules which affect the intracellular cascade.

25 The assay takes advantage of the ability of G protein coupled receptors to stimulate [³⁵S]GTPγS binding to membranes expressing the relevant receptors. The assay can, therefore, be used in the direct identification method to screen candidate compounds to constitutively activated G protein-coupled receptors. The assay is generic and has application to drug discovery at all G protein-coupled receptors.

30 The [³⁵S]GTPγS assay is incubated in 20 mM HEPES and between 1 and about 20mM MgCl₂ (this amount can be adjusted for optimization of results, although 20mM is preferred) pH 7.4, binding buffer with between about 0.3 and about 1.2 nM [³⁵S]GTPγS

(this amount can be adjusted for optimization of results, although 1.2 is preferred) and 12.5 to 75 μ g membrane protein (e.g., 293 cells expressing the G_s Fusion Protein; this amount can be adjusted for optimization) and 10 μ M GDP (this amount can be changed for optimization) for 1 hour. Wheatgerm agglutinin beads (25 μ l; Amersham) will then be added and the mixture incubated for another 30 minutes at room temperature. The tubes will be then centrifuged at 1500 x g for 5 minutes at room temperature and then counted in a scintillation counter.

2. Adenylyl Cyclase

A Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) designed for cell-based assays can be modified for use with crude plasma membranes. The Flash Plate wells can contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells can be quantitated by a direct competition for binding of radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the measurement of changes in cAMP levels in whole cells that express the receptors.

Transfected cells will be harvested approximately twenty four hours after transient transfection. Media will be carefully aspirated and discarded. Ten ml of PBS will gently be added to each dish of cells followed by careful aspiration. One ml of Sigma cell dissociation buffer and 3ml of PBS will be added to each plate. Cells will be pipetted off the plate and the cell suspension collected into a 50ml conical centrifuge tube. Cells will be centrifuged at room temperature at 1,100 rpm for 5 min. The cell pellet will be carefully re-suspended into an appropriate volume of PBS (about 3ml/plate). The cells will be then counted using a hemocytometer and additional PBS will be added to give the appropriate number of cells (to a final volume of about 50 μ l/well).

cAMP standards and Detection Buffer (comprising 1 μ Ci of tracer [¹²⁵I cAMP (50 μ l] to 11 ml Detection Buffer) will be prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer will be prepared fresh for screening and contained 50 μ l of Stimulation Buffer, 3 μ l of test compound (12 μ M final assay concentration) and 50 μ l cells, Assay Buffer will be stored on ice until utilized. The assay will be initiated by addition of 50 μ l of cAMP standards to appropriate wells followed by addition of 50 μ l of PBSA to wells H-11 and H12. Fifty μ l of Stimulation Buffer will be added to all wells. DMSO (or selected candidate compounds) will be added to appropriate

wells using a pin tool capable of dispensing 3 μ l of compound solution, with a final assay concentration of 12 μ M test compound and 100 μ l total assay volume. The cells will then be added to the wells and incubated for 60 min at room temperature. One hundred μ l of Detection Mix containing tracer cAMP will then be added to the wells. Plates will be
5 incubated for an additional 2 hours followed by counting in a Wallac MicroBeta™ scintillation counter. Values of cAMP/well will then be extrapolated from a standard cAMP curve which will be contained within each assay plate.

3. Cell-Based cAMP for G_i Coupled Target GPCRs

10 TSHR is a G_s coupled GPCR that causes the accumulation of cAMP upon activation. TSHR will be constitutively activated by mutating amino acid residue 623 (*i.e.*, changing an alanine residue to an isoleucine residue). A G_i coupled receptor is expected to inhibit adenylyl cyclase, and, therefore, decrease the level of cAMP production, which can make assessment of cAMP levels challenging. An effective technique for measuring the
15 decrease in production of cAMP as an indication of constitutive activation of a G_i coupled receptor can be accomplished by co-transfecting, most preferably, non-endogenous, constitutively activated TSHR (TSHR-A623I) (or an endogenous, constitutively active G_s coupled receptor) as a "signal enhancer" with a G_i linked target GPCR to establish a baseline level of cAMP. Upon creating a non-endogenous version of the G_i coupled
20 receptor, this non-endogenous version of the target GPCR is then co-transfected with the signal enhancer, and it is this material that can be used for screening. This approach will be utilized to effectively generate a signal when a cAMP assay is used; this approach is preferably used in the direct identification of candidate compounds against G_i coupled receptors. It is noted that for a G_i coupled GPCR, when this approach is used, an inverse
25 agonist of the target GPCR will increase the cAMP signal and an agonist will decrease the cAMP signal.

On day one, 2x10⁴ 293 cells/well will be plated out. On day two, two reaction tubes will be prepared (the proportions to follow for each tube are per plate): tube A will be prepared by mixing 2 μ g DNA of each receptor transfected into the mammalian cells, for a
30 total of 4 μ g DNA (*e.g.*, pCMV vector; pCMV vector with mutated THSR (TSHR-A623I); TSHR-A623I and GPCR, *etc.*) in 1.2ml serum free DMEM (Irvine Scientific, Irvine, CA); tube B will be prepared by mixing 120 μ l lipofectamine (Gibco BRL) in 1.2ml serum free

DMEM. Tubes A and B will then be admixed by inversion (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the "transfection mixture". Plated 293 cells will be washed with 1XPBS, followed by addition of 10ml serum free DMEM. 2.4ml of the transfection mixture will then be added
5 to the cells, followed by incubation for 4hrs at 37°C/5% CO₂. The transfection mixture will then be removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells will then be incubated at 37°C/5% CO₂. After 24hr incubation, cells will be harvested and utilized for analysis.

A Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A)
10 although designed for cell-based assays, can be modified for use with crude plasma membranes depending on the need of the skilled artisan. The Flash Plate wells will contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells can be quantified by a direct competition for binding of radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the
15 measurement of changes in cAMP levels in whole cells that express the receptors.

Transfected cells will be harvested approximately twenty four hours after transient transfection. Media will be carefully aspirated and discarded. Ten ml of PBS will be gently added to each dish of cells followed by careful aspiration. One ml of Sigma cell dissociation buffer and 3ml of PBS will be added to each plate. Cells will be
20 pipetted off the plate and the cell suspension will be collected into a 50ml conical centrifuge tube. Cells will be centrifuged at room temperature at 1,100 rpm for 5 min. The cell pellet will be carefully re-suspended into an appropriate volume of PBS (about 3ml/plate). The cells will then be counted using a hemocytometer and additional PBS is added to give the appropriate number of cells (to a final volume of about 50µl/well).

25 cAMP standards and Detection Buffer (comprising 1 µCi of tracer [¹²⁵I cAMP (50 µl] to 11 ml Detection Buffer) will be prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer should be prepared fresh for screening and contained 50µl of Stimulation Buffer, 3µl of test compound (12µM final assay concentration) and 50µl cells, Assay Buffer can be stored on ice until utilized. The assay
30 can be initiated by addition of 50µl of cAMP standards to appropriate wells followed by addition of 50µl of PBSA to wells H-11 and H12. Fifty µl of Stimulation Buffer will be added to all wells. Selected compounds (e.g., TSH) will be added to appropriate wells

using a pin tool capable of dispensing 3µl of compound solution, with a final assay concentration of 12µM test compound and 100µl total assay volume. The cells will then be added to the wells and incubated for 60 min at room temperature. One hundred µl of Detection Mix containing tracer cAMP will then be added to the wells. Plates will then be incubated additional 2 hours followed by counting in a Wallac MicroBeta scintillation counter. Values of cAMP/well will then be extrapolated from a standard cAMP curve which is contained within each assay plate.

4. Reporter-Based Assays

a. CRE-LUC Reporter Assay (G_s -associated receptors)

293 and 293T cells will be plated-out on 96 well plates at a density of 2×10^4 cells per well and will be transfected using Lipofectamine Reagent (BRL) the following day according to manufacturer instructions. A DNA/lipid mixture will be prepared for each 6-well transfection as follows: 260ng of plasmid DNA in 100µl of DMEM are gently mixed with 2µl of lipid in 100µl of DMEM (the 260ng of plasmid DNA consisted of 200ng of a 8xCRE-Luc reporter plasmid, 50ng of pCMV comprising endogenous receptor or non-endogenous receptor or pCMV alone, and 10ng of a GPRS expression plasmid (GPRS in pcDNA3 (Invitrogen)). The 8XCRE-Luc reporter plasmid is prepared as follows: vector SRIF-β-gal will be obtained by cloning the rat somatostatin promoter (-71/+51) at BglV-HindIII site in the pβgal-Basic Vector (Clontech). Eight (8) copies of cAMP response element will be obtained by PCR from an adenovirus template AdpCF126CCRE8 (*see, 7 Human Gene Therapy* 1883 (1996)) and cloned into the SRIF-β-gal vector at the Kpn-BglV site, resulting in the 8xCRE-β-gal reporter vector. The 8xCRE-Luc reporter plasmid will be generated by replacing the beta-galactosidase gene in the 8xCRE-β-gal reporter vector with the luciferase gene obtained from the pGL3-basic vector (Promega) at the HindIII-BamHI site. Following 30 min. incubation at room temperature, the DNA/lipid mixture will be diluted with 400 µl of DMEM and 100µl of the diluted mixture will be added to each well. One hundred µl of DMEM with 10% FCS will be added to each well after a 4hr incubation in a cell culture incubator. The following day the transfected cells will be changed with 200 µl/well of DMEM with 10% FCS. Eight hours later, the wells will be changed to 100 µl /well of DMEM without phenol red, after one wash with PBS. Luciferase activity will be measured the next day using the LucLite™ reporter gene assay kit (Packard) following manufacturer's

instructions and read on a 1450 MicroBeta™ scintillation and luminescence counter (Wallac).

b. AP1 reporter assay (G_q-associated receptors)

5 A method to detect G_q stimulation depends on the known property of G_q-dependent phospholipase C to cause the activation of genes containing AP1 elements in their promoter. A Pathdetect™ AP-1 cis-Reporting System (Stratagene, Catalogue # 219073) can be utilized following the protocol set forth above with respect to the CREB reporter assay, except that the components of the calcium phosphate precipitate were 410
10 ng pAP1-Luc, 80 ng pCMV-receptor expression plasmid, and 20 ng CMV-SEAP.

c. SRF-LUC Reporter Assay (G_q- associated receptors)

One method to detect G_q stimulation depends on the known property of G_q-dependent phospholipase C to cause the activation of genes containing serum response factors in their promoter. A Pathdetect™ SRF-Luc-Reporting System (Stratagene) can be
15 utilized to assay for G_q coupled activity in, *e.g.*, COS7 cells. Cells are transfected with the plasmid components of the system and the indicated expression plasmid encoding endogenous or non-endogenous GPCR using a Mammalian Transfection™ Kit (Stratagene, Catalogue #200285) according to the manufacturer's instructions. Briefly, 410 ng SRF-Luc, 80 ng pCMV-receptor expression plasmid and 20 ng CMV-SEAP
20 (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in transfection efficiency between samples) are combined in a calcium phosphate precipitate as per the manufacturer's instructions. Half of the precipitate is equally distributed between 3 wells in a 96-well plate, kept on the cells in a serum free media for 24 hours. The last 5
25 hours the cells are incubated with 1μM Angiotensin, where indicated. Cells are then lysed and assayed for luciferase activity using a Lucite™ Kit (Packard, Cat. # 6016911) and "Trilux 1450 Microbeta" liquid scintillation and luminescence counter (Wallac) as per the manufacturer's instructions. The data can be analyzed using GraphPad Prism™ 2.0a (GraphPad Software Inc.).

30 **d. Intracellular IP₃ Accumulation Assay (G_q-associated receptors)**

On day 1, cells comprising the receptors (endogenous and/or non-endogenous) are plated onto 24 well plates, usually 1×10^5 cells/well (although this number can be optimized). On day 2 cells are transfected by firstly mixing 0.25ug DNA in 50 μ l serum free DMEM/well and 2 μ l lipofectamine in 50 μ l serum free DMEM/well. The solutions are gently mixed and incubated for 15-30 min at room temperature. Cells are then washed with 0.5 ml PBS and 400 μ l of serum free media and then mixed with the transfection media and added to the cells. The cells are incubated for 3-4 hrs at $37^\circ\text{C}/5\%\text{CO}_2$ and then the transfection media is removed and replaced with 1ml/well of regular growth media. On day 3 the cells are labeled with ^3H -myo-inositol. Briefly, the media is removed and the cells are washed with 0.5 ml PBS. Then 0.5 ml inositol-free/serum free media (GIBCO BRL) are added/well with 0.25 μCi of ^3H -myo-inositol/well and the cells incubated for 16-18 hrs overnight at $37^\circ\text{C}/5\%\text{CO}_2$. On Day 4 the cells are washed with 0.5 ml PBS and 0.45 ml of assay medium is added containing inositol-free/serum free media 10 μM pargyline 10 mM lithium chloride or 0.4 ml of assay medium and 50 μ l of 10x ketanserin (ket) to final concentration of 10 μM . The cells are then incubated for 30 min at 37°C . The cells are then washed with 0.5 ml PBS and 200 μ l of fresh/ice cold stop solution (1M KOH; 18 mM Na-borate; 3.8 mM EDTA) is added to each well. The solution is kept on ice for 5-10 min (or until cells are lysed) and then neutralized by 200 μ l of fresh/ice cold neutralization solution (7.5 % HCL). The lysate is then transferred into 1.5 ml Eppendorf tubes and 1 ml of chloroform/methanol (1:2) is added/tube. The solution is vortexed for 15 sec and the upper phase is applied to a Biorad AG1-X8TM anion exchange resin (100-200 mesh). First, the resin is washed with water at 1:1.25 W/V and 0.9 ml of upper phase is loaded onto the column. The column is then washed with 10 ml of 5 mM myo-inositol and 10 ml of 5 mM Na-borate/60mM Na-formate. The inositol tris phosphates are eluted into scintillation vials containing 10 ml of scintillation cocktail with 2 ml of 0.1 M formic acid/ 1 M ammonium formate. The columns are regenerated by washing with 10 ml of 0.1 M formic acid/3M ammonium formate and rinsed twice with dd H₂O and stored at 4°C in water.

Reference is made to Figure 1. In Figure 1, 293 cells were transfected with G_q protein containing a six amino acid deletion, "G_q(del)"; G_q protein fused to a G_i protein, "G_q(del)/G_i"; endogenous RUP32; and RUP32 with G_q(del) ("RUP32+ G_q(del)/G_i"). The data indicate, based upon measuring IP₃ accumulation of RUP32 co-transfection of

$G_q(\text{del})/G_i$, that RUP32 does not endogenously couple to G_q protein. However when RUP32 was co-transfected with $G_q(\text{del})/G_i$ fusion protein, RUP32 was forced to couple to G_q protein. RUP27+ $G_q(\text{del})/G_i$ evidence about a nine (9) fold increase in IP3 accumulation when compared to endogenous RUP32. This data demonstrates that the $G_q(\text{del})/G_i$ Fusion Construct can be co-transfected with a GPCR and used to screen for agonists or inverse agonists.

Reference is made to Figure 2. In Figure 2, 293 cells were transfected with RUP35 and RUP36 receptor and compared to the control, pCMV. The data indicate that both RUP35 and RUP36 receptor are endogenously, constitutively active. RUP35 evidences about a six (6) fold increase in intracellular inositol phosphate accumulation when compared to pCMV and RUP36 evidences about a four (4) fold increase when compared to pCMV.

Example 5

FUSION PROTEIN PREPARATION

a. GPCR: G_s Fusion Construct

The design of the constitutively activated GPCR-G protein fusion construct can be accomplished as follows: both the 5' and 3' ends of the rat G protein $G_s\alpha$ (long form; Itoh, H. et al., 83 *PNAS* 3776 (1986)) is engineered to include a HindIII (5'-AAGCTT-3') sequence thereon. Following confirmation of the correct sequence (including the flanking HindIII sequences), the entire sequence is shuttled into pcDNA3.1(-) (Invitrogen, cat. no. V795-20) by subcloning using the HindIII restriction site of that vector. The correct orientation for the $G_s\alpha$ sequence will be determined after subcloning into pcDNA3.1(-). The modified pcDNA3.1(-) containing the rat $G_s\alpha$ gene at HindIII sequence is then verified; this vector will then be available as a "universal" $G_s\alpha$ protein vector. The pcDNA3.1(-) vector contains a variety of well-known restriction sites upstream of the HindIII site, thus beneficially providing the ability to insert, upstream of the G_s protein, the coding sequence of an endogenous, constitutively active GPCR. This same approach can be utilized to create other "universal" G protein vectors, and, of course, other commercially available or proprietary vectors known to the artisan can be utilized. In some embodiments, the important criteria is that the sequence for the GPCR be upstream and in-frame with that of the G protein.

Spacers in the restriction sites between the G protein and the GPCR are optional. The sense and anti-sense primers included the restriction sites for XbaI and EcoRV, respectively, such that spacers (attributed to the restriction sites) exist between the G protein and the GPCR.

5 PCR will then be utilized to secure the respective receptor sequences for fusion within the $G_s\alpha$ universal vector disclosed above, using the following protocol for each: 100ng cDNA for GPCR will be added to separate tubes containing 2 μ l of each primer (sense and anti-sense), 3 μ l of 10mM dNTPs, 10 μ l of 10XTaqPlus™ Precision buffer, 1 μ l of TaqPlus™ Precision polymerase (Stratagene: #600211), and 80 μ l of water. Reaction
10 temperatures and cycle times for the GPCR will be as follows with cycle steps 2 through 4 were repeated 35 times: 94°C for 1 min; 94°C for 30 seconds; 62°C for 20 sec; 72°C 1 min 40sec; and 72°C 5 min. PCR products will be run on a 1% agarose gel and then purified. The purified products will be digested with XbaI and EcoRV and the desired inserts purified and ligated into the G_s universal vector at the respective restriction sites. The
15 positive clones will be isolated following transformation and determined by restriction enzyme digestion; expression using 293 cells will be accomplished following the protocol set forth *infra*. Each positive clone for GPCR- G_s Fusion Protein will be sequenced to verify correctness.

b. G_q (6 amino acid deletion)/ G_i Fusion Construct

20 The design of a G_q (del)/ G_i fusion construct was accomplished as follows: the N-terminal six (6) amino acids (amino acids 2 through 7), having the sequence of TLESIM (SEQ.ID.NO.:47) $G\alpha q$ -subunit was deleted and the C-terminal five (5) amino acids, having the sequence EYNLV (SEQ.ID.NO.:48) was replaced with the corresponding amino acids of the $G\alpha i$ Protein, having the sequence DCGLF (SEQ.ID.NO.:49). This fusion construct
25 was obtained by PCR using the following primers:
5'-gatcAAGCTTCCATGGCGTGCTGCCTGAGCGAGG-3' (SEQ.ID.NO.:50) and
5'-gatcGGATCCTTAGAACAGGCCGCGAGTCCTTCAGGTTTCAGCTGCAGGATGGTG-3'
(SEQ.ID.NO.:51) and Plasmid 63313 which contains the mouse $G\alpha q$ -wild type version with a hemagglutinin tag as template. Nucleotides in lower caps are included as spacers.

30 TaqPlus® Precision DNA polymerase (Stratagene) was utilized for the amplification by the following cycles, with steps 2 through 4 repeated 35 times: 95°C for 2 min; 95°C for 20 sec; 56°C for 20 sec; 72°C for 2 min; and 72°C for 7 min. The PCR

AREN-0309**PATENT**

product will be cloned into a pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator kit (P.E. Biosystems). Inserts from a TOPO clone containing the sequence of the fusion construct will be shuttled into the expression vector pcDNA3.1(+) at the HindIII/BamHI site by a 2 step cloning process.

5 Example 6**TISSUE DISTRIBUTION OF THE DISCLOSED HUMAN GPCRS: RT-PCR**

RT-PCR was applied to confirm the expression and to determine the tissue distribution of several novel human GPCRs. Oligonucleotides utilized were GPCR-specific and the human multiple tissue cDNA panels (MTC, Clontech) as templates. Taq DNA polymerase (Stratagene) were utilized for the amplification in a 40µl reaction according to the manufacturer's instructions. Twenty µl of the reaction will be loaded on a 1.5% agarose gel to analyze the RT-PCR products. Table E, below, lists the receptors, the cycle conditions and the primers utilized, and also lists exemplary diseases/disorders linked to the receptors.

TABLE E

Receptor Identifier	Cycle Conditions Min ('), Sec (") Cycles 2-4 repeated 35 times	5' Primer (SEQ.ID.NO.)	3' Primer (SEQ.ID.NO.)	DNA Fragment	Tissue Expression
hRUP28	94°C for 5 min; 94°C for 30 sec; 58°C for 30 sec, 72°C for 1 min, and 72°C for 7 min	GTCCTCACT GGTGGCCAT GTACTCC (52)	CTGCGTCCAC CAGAGTCAC GTCTCC (53)	710bp	heart; kidney; liver; lung and pancreas
hRUP29	94°C for 5 min; 94°C for 30 sec; 58°C for 30 sec, 72°C for 1 min,	CTTGATGTT TGGGCTGCC CTTCTGC (54)	GTTTGTGGCT AACGGCACA AAACACAAT TCC (55)	690bp	leukocyte and ovary

AREN-0309
PATENT

	and 72°C for 7 min				
hRUP30	94°C for 2 min; 94°C for 15 sec; 58°C for 20 sec, 72°C for 1 min, and 72°C for 10 min	CTGCTCACG GTTGACCGC TACACTGC (56)	GTGGCCATG AGCCACCCT GAGCTCC (57)	690bp	pancreas
hRUP31	95°C for 4 min; 95°C for 1 min; 52°C for 30 sec, 72°C for 1 min, and 72°C for 7 min	CTTCTTCTCC GACGTCAAG G (58)	CCAAATCA GTGTGCAA ATCG (59)	516bp	colon, lung, pancreas, thymus; cerebral cortex, hippocampus of brain, and fat cells
hRUP32	95°C for 4 min; 95°C for 1 min; 52°C for 30 sec, 72°C for 1 min, and 72°C for 7 min	TGAATGGGT CCTGTGTGA AA (60)	CAACGGTCT GACAACCTC CT (61)	527bp	thymus
hRUP34	95°C for 4 min; 95°C for 1 min; 52°C for 30 sec, 72°C for 1 min, and 72°C for 7 min	TTGCTGTGAT GTGGCATT G (62)	CAGGAAGCC CATAAAGGC ATCAA (63)	534bp	peripheral blood leukocyte ("PBL"), prostate and kidney
hRUP35	95°C for 4 min; 95°C for 1 min; 52°C for 30 sec, 72°C for 1 min, and 72°C for 7 min	ACATCACCT GCTTCTGA CC (64)	CCAGCATCTT GATGCAGTG T (65)	557bp	thalamus
hRUP37	95°C for 4 min; 95°C for 1 min; 52°C for 30 sec, 72°C for 1 min,	CCATCTCCA AAATCCTCA GTC (66)	GCTGTTAAG AGCGGACAG GAAA (67)	517bp	testis, cerebral cortex and hippocampus

	and 72°C for 7 min				
--	-----------------------	--	--	--	--

Diseases and disorders related to receptors located in these tissues or regions include, but are not limited to, cardiac disorders and diseases (e.g. thrombosis, myocardial infarction; atherosclerosis; cardiomyopathies); kidney disease/disorders (e.g., renal failure; renal tubular acidosis; renal glycosuria; nephrogenic diabetes insipidus; cystinuria; polycystic kidney disease); eosinophilia; leukocytosis; leukopenia; ovarian cancer; sexual dysfunction; polycystic ovarian syndrome; pancreatitis and pancreatic cancer; irritable bowel syndrome; colon cancer; Crohn's disease; ulcerative colitis; diverticulitis; Chronic Obstructive Pulmonary Disease (COPD); Cystic Fibrosis; pneumonia; pulmonary hypertension; tuberculosis and lung cancer; Parkinson's disease; movement disorders and ataxias; learning and memory disorders; eating disorders (e.g., anorexia; bulimia, etc.); obesity; cancers; thymoma; myasthenia gravis; circulatory disorders; prostate cancer; prostatitis; kidney disease/disorders (e.g., renal failure; renal tubular acidosis; renal glycosuria; nephrogenic diabetes insipidus; cystinuria; polycystic kidney disease); sensorimotor processing and arousal disorders; obsessive-compulsive disorders; testicular cancer; priapism; prostatitis; hernia; endocrine disorders; sexual dysfunction; allergies; depression; psychotic disorders; migraine; reflux; schizophrenia; ulcers; bronchospasm; epilepsy; prostatic hypertrophy; anxiety; rhinitis; angina; and glaucoma. Accordingly, the methods of the present invention may also be useful in the diagnosis and/or treatment of these and other diseases and disorders.

Example 7

Protocol: Direct Identification of Inverse Agonists and Agonists

A. [³⁵S]GTP γ S Assay

Although endogenous, constitutively active GPCRs have been used for the direct identification of candidate compounds as, e.g., inverse agonists, for reasons that are not altogether understood, intra-assay variation can become exacerbated. In some embodiments a GPCR Fusion Protein, as disclosed above, is also utilized with a non-endogenous, constitutively activated GPCR. When such a protein is used, intra-assay variation appears to be substantially stabilized, whereby an effective signal-to-noise ratio is

obtained. This has the beneficial result of allowing for a more robust identification of candidate compounds. Thus, in some embodiments it is preferred that for direct identification, a GPCR Fusion Protein be used and that when utilized, the following assay protocols be utilized.

5 1. **Membrane Preparation**

Membranes comprising the constitutively active orphan GPCR Fusion Protein of interest and for use in the direct identification of candidate compounds as inverse agonists or agonists are preferably prepared as follows:

 a. **Materials**

10 “Membrane Scrape Buffer” is comprised of 20mM HEPES and 10mM EDTA, pH 7.4; “Membrane Wash Buffer” is comprised of 20 mM HEPES and 0.1 mM EDTA, pH 7.4; “Binding Buffer” is comprised of 20mM HEPES, 100 mM NaCl, and 10 mM MgCl₂, pH 7.4

 b. **Procedure**

15 All materials will be kept on ice throughout the procedure. Firstly, the media will be aspirated from a confluent monolayer of cells, followed by rinse with 10ml cold PBS, followed by aspiration. Thereafter, 5ml of Membrane Scrape Buffer will be added to scrape cells; this will be followed by transfer of cellular extract into 50ml centrifuge tubes (centrifuged at 20,000 rpm for 17 minutes at 4°C). Thereafter, the supernatant will be
20 aspirated and the pellet will be resuspended in 30ml Membrane Wash Buffer followed by centrifugation at 20,000 rpm for 17 minutes at 4°C. The supernatant will then be aspirated and the pellet resuspended in Binding Buffer. The resuspended pellet will then be homogenized using a Brinkman Polytron™ homogenizer (15-20 second bursts until the material is in suspension). This is referred to herein as “Membrane Protein”.

25 2. **Bradford Protein Assay**

Following the homogenization, protein concentration of the membranes will be determined, for example, using the Bradford Protein Assay (protein can be diluted to about 1.5mg/ml, aliquoted and frozen (-80°C) for later use; when frozen, protocol for use will be as follows: on the day of the assay, frozen Membrane Protein is thawed at room
30 temperature, followed by vortex and then homogenized with a Polytron at about 12 x 1,000 rpm for about 5-10 seconds; it was noted that for multiple preparations, the homogenizer is thoroughly cleaned between homogenization of different preparations).

a. Materials

Binding Buffer (as discussed above); Bradford Dye Reagent; Bradford Protein Standard will be utilized, following manufacturer instructions (Biorad, cat. no. 500-0006).

5

b. Procedure

Duplicate tubes will be prepared, one including the membrane, and one as a control "blank". Each contains 800 μ l Binding Buffer. Thereafter, 10 μ l of Bradford Protein Standard (1mg/ml) will be added to each tube, and 10 μ l of membrane Protein will then be added to just one tube (not the blank). Thereafter, 200 μ l of Bradford Dye
10 Reagent will be added to each tube, followed by vortexing. After five minutes, the tubes will be re-vortexed and the material therein will be transferred to cuvettes. The cuvettes will then be read using a CECIL 3041 spectrophotometer, at wavelength 595.

3. Direct Identification Assay**a. Materials**

15

GDP Buffer consisted of 37.5 ml Binding Buffer and 2mg GDP (Sigma, cat. no. G-7127), followed by a series of dilutions in Binding Buffer to obtain 0.2 μ M GDP (final concentration of GDP in each well was 0.1 μ M GDP); each well comprising a candidate compound, has a final volume of 200 μ l consisting of 100 μ l GDP Buffer (final concentration, 0.1 μ M GDP), 50 μ l Membrane Protein in Binding Buffer, and 50 μ l
20 [35 S]GTP γ S (0.6 nM) in Binding Buffer (2.5 μ l [35 S]GTP γ S per 10ml Binding Buffer).

b. Procedure

Candidate compounds will be preferably screened using a 96-well plate format (these can be frozen at -80°C). Membrane Protein (or membranes with expression vector excluding the GPCR Fusion Protein, as control), will be homogenized briefly until in
25 suspension. Protein concentration will then be determined using, for example, the Bradford Protein Assay set forth above. Membrane Protein (and controls) will then be diluted to 0.25mg/ml in Binding Buffer (final assay concentration, 12.5 μ g/well). Thereafter, 100 μ l GDP Buffer is added to each well of a Wallac ScintistripTM (Wallac). A 5 μ l pin-tool will then be used to transfer 5 μ l of a candidate compound into such well (*i.e.*, 5 μ l in total assay
30 volume of 200 μ l is a 1:40 ratio such that the final screening concentration of the candidate compound is 10 μ M). Again, to avoid contamination, after each transfer step the pin tool is rinsed in three reservoirs comprising water (1X), ethanol (1X) and water (2X) – excess

liquid is shaken from the tool after each rinse and the tool is dried with paper and Kim wipes. Thereafter, 50 μ l of Membrane Protein will be added to each well (a control well comprising membranes without the GPCR Fusion Protein was also utilized), and pre-incubated for 5-10 minutes at room temperature. Thereafter, 50 μ l of [35 S]GTP γ S, (0.6 nM) in Binding Buffer will be added to each well, followed by incubation on a shaker for 60 minutes at room temperature (again, in this example, plates were covered with foil). The assay will be stopped by spinning the plates at 4000 RPM for 15 minutes at 22°C. The plates will then be aspirated with an 8 channel manifold and sealed with plate covers. The plates will then be read on a Wallac 1450 using setting "Prot. #37" (as per manufacturer's instructions).

B. Cyclic AMP Assay

Another assay approach to directly identify candidate compound will be accomplished utilizing a cyclase-based assay. In addition to direct identification, this assay approach can be utilized as an independent approach to provide confirmation of the results from the [35 S]GTP γ S approach as set forth above.

A modified Flash PlateTM Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) will be preferably utilized for direct identification of candidate compounds as inverse agonists and agonists to GPCRs in accordance with the following protocol.

Transfected cells will be harvested approximately three days after transfection. Membranes will be prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂. Homogenization will be performed on ice using a Brinkman PolytronTM for approximately 10 seconds. The resulting homogenate will be centrifuged at 49,000 X g for 15 minutes at 4°C. The resulting pellet will then be resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet will then be stored at -80°C until utilized. On the day of direct identification screening, the membrane pellet will slowly be thawed at room temperature, resuspended in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂, to yield a final protein concentration of 0.60mg/ml (the resuspended membranes will be placed on ice until use).

cAMP standards and Detection Buffer (comprising 2 μ Ci of tracer [125 I cAMP (100 μ l] to 11 ml Detection Buffer) will be prepared and maintained in accordance with the

AREN-0309**PATENT**

manufacturer's instructions. Assay Buffer will be prepared fresh for screening and contain 20mM HEPES, pH 7.4, 10mM MgCl₂, 20mM phosphocreatine (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 μ M GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer will be stored on ice until utilized.

5 Candidate compounds identified as per above (if frozen, thawed at room temperature) will be added, preferably, to 96-well plate wells (3 μ l/well; 12 μ M final assay concentration), together with 40 μ l Membrane Protein (30 μ g/well) and 50 μ l of Assay Buffer. This admixture will be incubated for 30 minutes at room temperature, with gentle shaking.

10 Following the incubation, 100 μ l of Detection Buffer will be added to each well, followed by incubation for 2-24 hours. Plates will then be counted in a Wallac MicroBeta™ plate reader using "Prot. #31" (as per manufacturer instructions).

C. Melanophore Screening Assay

A method for identifying candidate agonists or inverse agonists for a GPCR can be
15 performed by introducing tests cells of a pigment cell line capable of dispersing or aggregating their pigment in response to a specific stimulus and expressing an exogenous clone coding for the GPCR. A stimulant, *e.g.*, light, sets an initial state of pigment disposition wherein the pigment is aggregated within the test cells if activation of the GPCR induces pigment dispersion. However, stimulating the cell with a stimulant to set an initial
20 state of pigment disposition wherein the pigment is dispersed if activation of the GPCR induces pigment aggregation. The tests cells are then contacted with chemical compounds, and it is determined whether the pigment disposition in the cells changed from the initial state of pigment disposition. Dispersion of pigments cells due to the candidate compound coupling to the GPCR will appear dark on a petri dish, while aggregation of pigments cells
25 will appear light.

Materials and methods will be followed according to the disclosure of U.S. Patent Number 5,462,856 and U.S. Patent Number 6,051,386, each of which are incorporated by reference.

Although a variety of expression vectors are available to those in the art, for
30 purposes of utilization for both the endogenous and non-endogenous human GPCRs, in some embodiments it is preferred that the vector utilized be pCMV. This vector was deposited with the American Type Culture Collection (ATCC) on October 13, 1998 (10801

AREN-0309**PATENT**

University Blvd., Manassas, VA 20110-2209 USA) under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure. The DNA was tested by the ATCC and determined to be viable. The ATCC has assigned the following deposit number to pCMV: ATCC #203351.

- 5 References cited throughout this patent document, including co-pending and related patent applications, unless otherwise indicated, are fully incorporated herein by reference. Modifications and extension of the disclosed inventions that are within the purview of the skilled artisan are encompassed within the above disclosure and the claims that follow.

CLAIMS

What is claimed is:

1. A G protein-coupled receptor encoded by an amino acid sequence of
5 SEQ.ID.NO.:2.
2. A non-endogenous, constitutively activated version of the G protein-coupled
receptor of claim 1.
3. A plasmid comprising a vector and the cDNA of SEQ.ID.NO.:1.
4. A host cell comprising the plasmid of claim 3.
- 10 5. A G protein-coupled receptor encoded by an amino acid sequence of
SEQ.ID.NO.:4.
6. A non-endogenous, constitutively activated version of the G protein-coupled
receptor of claim 5.
7. A plasmid comprising a vector and the cDNA of SEQ.ID.NO.:3.
- 15 8. A host cell comprising the plasmid of claim 7.
9. A G protein-coupled receptor encoded by an amino acid sequence of
SEQ.ID.NO.:6.
10. A non-endogenous, constitutively activated version of the G protein-coupled
receptor of claim 9.
- 20 11. A plasmid comprising a vector and the cDNA of SEQ.ID.NO.:5.
12. A host cell comprising the plasmid of claim 11.
13. A G protein-coupled receptor encoded by an amino acid sequence of
SEQ.ID.NO.:8.
14. A non-endogenous, constitutively activated version of the G protein-coupled
25 receptor of claim 13.
15. A plasmid comprising a vector and the cDNA of SEQ.ID.NO.:7.
16. A host cell comprising the plasmid of claim 15.
17. A G protein-coupled receptor encoded by an amino acid sequence of
SEQ.ID.NO.:10.
- 30 18. A non-endogenous, constitutively activated version of the G protein-coupled
receptor of claim 17.
19. A plasmid comprising a vector and the cDNA of SEQ.ID.NO.:9.

20. A host cell comprising the plasmid of claim 19.
21. A G protein-coupled receptor encoded by an amino acid sequence of
SEQ.ID.NO.:12.
22. A non-endogenous, constitutively activated version of the G protein-coupled
5 receptor of claim 21.
23. A plasmid comprising a vector and the cDNA of SEQ.ID.NO.:11.
24. A host cell comprising the plasmid of claim 23.
25. A G protein-coupled receptor encoded by an amino acid sequence of
SEQ.ID.NO.:14.
- 10 26. A non-endogenous, constitutively activated version of the G protein-coupled
receptor of claim 25.
27. A plasmid comprising a vector and the cDNA of SEQ.ID.NO.:13.
28. A host cell comprising the plasmid of claim 27.
29. A G protein-coupled receptor encoded by an amino acid sequence of
15 SEQ.ID.NO.:16.
30. A non-endogenous, constitutively activated version of the G protein-coupled
receptor of claim 29.
31. A plasmid comprising a vector and the cDNA of SEQ.ID.NO.:15.
32. A host cell comprising the plasmid of claim 31.
- 20 33. A G protein-coupled receptor encoded by an amino acid sequence of
SEQ.ID.NO.:18.
34. A non-endogenous, constitutively activated version of the G protein-coupled
receptor of claim 33.
35. A plasmid comprising a vector and the cDNA of SEQ.ID.NO.:17.
- 25 36. A host cell comprising the plasmid of claim 35.
37. A G protein-coupled receptor encoded by an amino acid sequence of
SEQ.ID.NO.:20.
38. A non-endogenous, constitutively activated version of the G protein-coupled
receptor of claim 37.
- 30 39. A plasmid comprising a vector and the cDNA of SE.ID.NO.:19.
40. A host cell comprising the plasmid of claim 39.

Figure 1

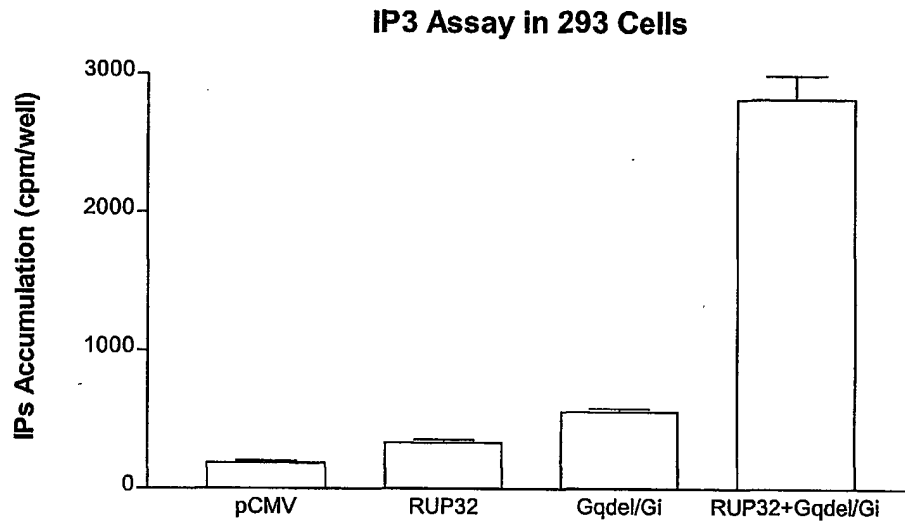
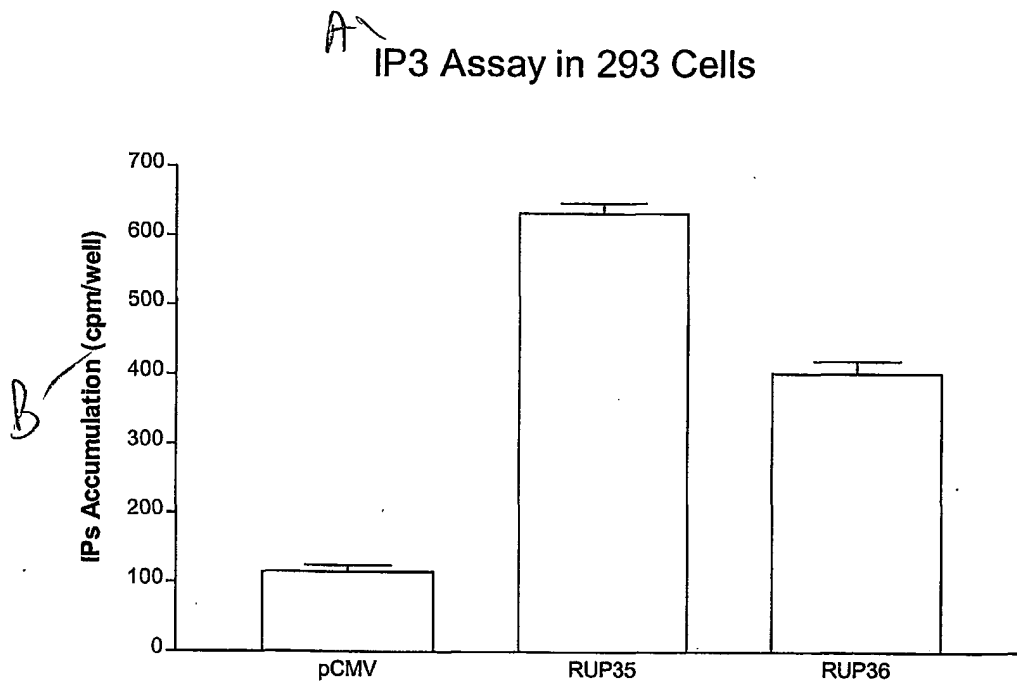


Figure 2



SEQUENCE LISTING

<110> Arena Pharmaceuticals, Inc.

<120> Endogenous And Non-Endogenous, Constitutively Activated Human G Protein-Coupled Receptors

<130> AREN-0309

<150> 09/170,496
<151> 1998-10-13

<150> PCT/US99/23938
<151> 1998-10-13

<150> 60/253,404
<151> 2000-11-27

<150> 60/255,366
<151> 2000-12-12

<150> 60/270,286
<151> 2001-02-20

<150> 60/282,365
<151> 2001-04-06

<150> 60/270,266
<151> 2001-02-20

<150> 60/282,032
<151> 2001-04-06

<150> 60/282,358
<151> 2001-04-06

<150> 60/282,356
<151> 2001-04-06

<150> 60/290,917
<151> 2001-05-14

<150> 60/309,208
<151> 2001-07-31

<160> 67

<170> PatentIn version 3.1

<210> 1
<211> 1002
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 1
atgtggagct gcagctggtt caacggcaca gggctggtgg aggagctgcc tgcctgccag 60
gacctgcagc tggggctgtc actgttgtcg ctgctgggcc tgggtggtgg cgtgccagtg 120
ggcctgtgct acaacgccct gctggtgctg gccaacctac acagcaaggc cagcatgacc 180
atgccggagc tgtactttgt caacatggca gtggcaggcc tgggtgctcag cgccctggcc 240

cctgtgcacc tgctcgcccc cccgagctcc cgggtgggcgc tgtggagtgt gggcggcgaa 300
 gtccacgtgg cactgcagat ccccttcaat gtgtcctcac tgggtggccat gtactccacc 360
 gccctgctga gcctcgacca ctacatcgag cgtgcactgc cgcggaccta catggccagc 420
 gtgtacaaca cgcggcacgt gtgcggcttc gtgtggggtg gcgcgctgct gaccagcttc 480
 tcctcgctgc tcttctacat ctgcagccat gtgtccaccc gcgcgctaga gtgcgccaag 540
 atgcagaacg cagaagctgc cgacgccacg ctggtgttca tcggctacgt ggtgccagca 600
 ctggccaccc tctacgcgct ggtgtactc tcccgcgtcc gcagggagga cagccccctg 660
 gaccgggaca cgggcccgtt ggagccctcg gcacacagga tgctggtggc caccgtgtgc 720
 acgcagtttg ggtcttgga gccacactat ctgatcctgc tggggcacac ggtcatcatc 780
 tcgcgaggga agcccgtaga cgcacactac ctggggctac tgcactttgt gaaggatttc 840
 tccaaactcc tggccttctc cagcagcttt gtgacaccac ttctctaccg ctacatgaac 900
 cagagcttcc ccagcaagct ccaacggctg atgaaaaagc tgcctgcgg ggaccggcac 960
 tgctccccgg accacatggg ggtgcagcag gtgctggcgt ag 1002

<210> 2
 <211> 333
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 2

Met Trp Ser Cys Ser Trp Phe Asn Gly Thr Gly Leu Val Glu Glu Leu
 1 5 10 15

Pro Ala Cys Gln Asp Leu Gln Leu Gly Leu Ser Leu Leu Ser Leu Leu
 20 25 30

Gly Leu Val Val Gly Val Pro Val Gly Leu Cys Tyr Asn Ala Leu Leu
 35 40 45

Val Leu Ala Asn Leu His Ser Lys Ala Ser Met Thr Met Pro Asp Val
 50 55 60

Tyr Phe Val Asn Met Ala Val Ala Gly Leu Val Leu Ser Ala Leu Ala
 65 70 75 80

Pro Val His Leu Leu Gly Pro Pro Ser Ser Arg Trp Ala Leu Trp Ser
 85 90 95

Val Gly Gly Glu Val His Val Ala Leu Gln Ile Pro Phe Asn Val Ser
 100 105 110

Ser Leu Val Ala Met Tyr Ser Thr Ala Leu Leu Ser Leu Asp His Tyr

115	120	125
Ile Glu Arg Ala Leu Pro Arg Thr Tyr Met Ala Ser Val Tyr Asn Thr 130 135 140		
Arg His Val Cys Gly Phe Val Trp Gly Gly Ala Leu Leu Thr Ser Phe 145 150 155 160		
Ser Ser Leu Leu Phe Tyr Ile Cys Ser His Val Ser Thr Arg Ala Leu 165 170 175		
Glu Cys Ala Lys Met Gln Asn Ala Glu Ala Ala Asp Ala Thr Leu Val 180 185 190		
Phe Ile Gly Tyr Val Val Pro Ala Leu Ala Thr Leu Tyr Ala Leu Val 195 200 205		
Leu Leu Ser Arg Val Arg Arg Glu Asp Thr Pro Leu Asp Arg Asp Thr 210 215 220		
Gly Arg Leu Glu Pro Ser Ala His Arg Leu Leu Val Ala Thr Val Cys 225 230 235 240		
Thr Gln Phe Gly Leu Trp Thr Pro His Tyr Leu Ile Leu Leu Gly His 245 250 255		
Thr Val Ile Ile Ser Arg Gly Lys Pro Val Asp Ala His Tyr Leu Gly 260 265 270		
Leu Leu His Phe Val Lys Asp Phe Ser Lys Leu Leu Ala Phe Ser Ser 275 280 285		
Ser Phe Val Thr Pro Leu Leu Tyr Arg Tyr Met Asn Gln Ser Phe Pro 290 295 300		
Ser Lys Leu Gln Arg Leu Met Lys Lys Leu Pro Cys Gly Asp Arg His 305 310 315 320		
Cys Ser Pro Asp His Met Gly Val Gln Gln Val Leu Ala 325 330		

<210> 3
 <211> 918
 <212> DNA
 <213> Homo sapiens

<400> 3 atgcctggcc acaatacctc caggaattcc tcttgcgatc ctatagtac accccactta	60
atcagcctct acttcatagt gcttattggc gggctggtgg gtgtcatttc cattcttttc	120
ctcctggtga aaatgaacac ccggtcagtg accaccatgg cggtcattaa cttggtggtg	180

gtccacagcg tttttctgct gacagtgcc tttcgcttga cctacctcat caagaagact 240
 tggatgtttg ggctgccctt ctgcaaattt gtgagtgcc tgctgcacat ccacatgtac 300
 ctcacgttcc tattctatgt ggtgatcctg gtcaccagat acctcatctt cttcaagtgc 360
 aaagacaaag tggaaattcta cagaaaactg catgctgtgg ctgccagtgc tggcatgtgg 420
 acgctggtga ttgtcattgt ggtaccctg gttgtctccc ggtatggaat ccatgaggaa 480
 tacaatgagg agcactgttt taaatttcac aaagagcttg cttacacata tgtgaaaatc 540
 atcaactata tgatagtcac ttttgcata gccgttgctg tgattctgtt ggtcttccag 600
 gtcttcatca ttatgttgat ggtgcagaag ctacgccact ctttactatc ccaccaggag 660
 ttctgggctc agctgaaaaa cctatTTTTT ataggggtca tccttggttg ttctcttccc 720
 taccagttct ttaggatcta ttacttgaat gttgtgacgc attccaatgc ctgtaacagc 780
 aaggttgcat ttataacga aatcttcttg agtgaacag caattagctg ctatgatttg 840
 cttctctttg tctttggggg aagccattgg ttttaagcaa agataattgg cttatggaat 900
 tgtgttttgt gccgttag 918

<210> 4
 <211> 305
 <212> PRT
 <213> Homo sapiens

<400> 4

Met Pro Gly His Asn Thr Ser Arg Asn Ser Ser Cys Asp Pro Ile Val
 1 5 10 15

Thr Pro His Leu Ile Ser Leu Tyr Phe Ile Val Leu Ile Gly Gly Leu
 20 25 30

Val Gly Val Ile Ser Ile Leu Phe Leu Leu Val Lys Met Asn Thr Arg
 35 40 45

Ser Val Thr Thr Met Ala Val Ile Asn Leu Val Val Val His Ser Val
 50 55 60

Phe Leu Leu Thr Val Pro Phe Arg Leu Thr Tyr Leu Ile Lys Lys Thr
 65 70 75 80

Trp Met Phe Gly Leu Pro Phe Cys Lys Phe Val Ser Ala Met Leu His
 85 90 95

Ile His Met Tyr Leu Thr Phe Leu Phe Tyr Val Val Ile Leu Val Thr
 100 105 110

Arg Tyr Leu Ile Phe Phe Lys Cys Lys Asp Lys Val Glu Phe Tyr Arg
 115 120 125

Lys Leu His Ala Val Ala Ala Ser Ala Gly Met Trp Thr Leu Val Ile
 130 135 140

Val Ile Val Val Pro Leu Val Val Ser Arg Tyr Gly Ile His Glu Glu
 145 150 155 160

Tyr Asn Glu Glu His Cys Phe Lys Phe His Lys Glu Leu Ala Tyr Thr
 165 170 175

Tyr Val Lys Ile Ile Asn Tyr Met Ile Val Ile Phe Val Ile Ala Val
 180 185 190

Ala Val Ile Leu Leu Val Phe Gln Val Phe Ile Ile Met Leu Met Val
 195 200 205

Gln Lys Leu Arg His Ser Leu Leu Ser His Gln Glu Phe Trp Ala Gln
 210 215 220

Leu Lys Asn Leu Phe Phe Ile Gly Val Ile Leu Val Cys Phe Leu Pro
 225 230 235 240

Tyr Gln Phe Phe Arg Ile Tyr Tyr Leu Asn Val Val Thr His Ser Asn
 245 250 255

Ala Cys Asn Ser Lys Val Ala Phe Tyr Asn Glu Ile Phe Leu Ser Val
 260 265 270

Thr Ala Ile Ser Cys Tyr Asp Leu Leu Leu Phe Val Phe Gly Gly Ser
 275 280 285

His Trp Phe Lys Gln Lys Ile Ile Gly Leu Trp Asn Cys Val Leu Cys
 290 295 300

Arg
 305

<210> 5
 <211> 1125
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 5
 atgctgacag ggagctgcgg ggaccctcag aaaaagccac aggtgaccca ggactcaggg 60
 cccagagca tggggcttga gggacgagag acagctggcc agccacgagt gaccctgctg 120
 cccacgcca acgtcagcgg gctgagccag gagtttgaaa gccactggcc agagatcgca 180
 gagaggtccc cgtgtgtggc tggcgtcatc cctgtcatct actacagtgt cctgctgggc 240

ttggggctgc ctgtcagcct cctgaccgca gtggccctgg cgcgccttgc caccaggacc 300
 aggaggccct cctactacta ccttctggcg ctacacagcct cggatatcat catccaggtg 360
 gtcacgtgtg tcgcgggctt cctcctgcag ggagcagtgc tggcccgcga ggtgcccag 420
 gctgtggtgc gcacggccaa catcctggag tttgtgcca accacgcctc agtctggatc 480
 gccatcctgc tcacggttga ccgctacact gccctgtgcc accccctgca ccatcggggc 540
 gcctcgtccc caggccggac ccgcccggcc attgtgtgtg tcctgagtgc tgccctgttg 600
 accggcatcc ccttctactg gtggctggac atgtggagag acaccgactc acccagaaca 660
 ctggacgagg tcctcaagtg ggctcactgt ctactgtct atttcatccc ttgtggcgtg 720
 ttctgtgtca ccaactcggc catcatccac cggctacgga ggaggggccc gagtgggctg 780
 cagccccggg tgggcaagag cacagccatc ctctgggca tcaccacact gttcacccctc 840
 ctgtgggcgc cccgggtctt cgtcatgctc taccacatgt acgtggcccc tgtccaccgg 900
 gactggaggg tcacactggc cttggatgtg gccaatatgg tggccatgct ccacacggca 960
 gccaaacttcg gcctctactg ctttgtcagc aagactttcc gggccactgt ccgacaggtc 1020
 atccacgatg cctacctgcc ctgcactttg gcacacagc cagagggcat ggcggcgaag 1080
 cctgtgatgg agcctccggg actccccaca ggggcagaag tgtag 1125

<210> 6
 <211> 374
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 6

Met Leu Thr Gly Ser Cys Gly Asp Pro Gln Lys Lys Pro Gln Val Thr
1 5 10 15

Gln Asp Ser Gly Pro Gln Ser Met Gly Leu Glu Gly Arg Glu Thr Ala
20 25 30

Gly Gln Pro Arg Val Thr Leu Leu Pro Thr Pro Asn Val Ser Gly Leu
35 40 45

Ser Gln Glu Phe Glu Ser His Trp Pro Glu Ile Ala Glu Arg Ser Pro
50 55 60

Cys Val Ala Gly Val Ile Pro Val Ile Tyr Tyr Ser Val Leu Leu Gly
65 70 75 80

Leu Gly Leu Pro Val Ser Leu Leu Thr Ala Val Ala Leu Ala Arg Leu
85 90 95

Ala Thr Arg Thr Arg Arg Pro Ser Tyr Tyr Tyr Leu Leu Ala Leu Thr

100	105	110
Ala Ser Asp Ile Ile Ile Gln Val Val Ile Val Phe Ala Gly Phe Leu 115 120 125		
Leu Gln Gly Ala Val Leu Ala Arg Gln Val Pro Gln Ala Val Val Arg 130 135 140		
Thr Ala Asn Ile Leu Glu Phe Ala Ala Asn His Ala Ser Val Trp Ile 145 150 155 160		
Ala Ile Leu Leu Thr Val Asp Arg Tyr Thr Ala Leu Cys His Pro Leu 165 170 175		
His His Arg Ala Ala Ser Ser Pro Gly Arg Thr Arg Arg Ala Ile Ala 180 185 190		
Ala Val Leu Ser Ala Ala Leu Leu Thr Gly Ile Pro Phe Tyr Trp Trp 195 200 205		
Leu Asp Met Trp Arg Asp Thr Asp Ser Pro Arg Thr Leu Asp Glu Val 210 215 220		
Leu Lys Trp Ala His Cys Leu Thr Val Tyr Phe Ile Pro Cys Gly Val 225 230 235 240		
Phe Leu Val Thr Asn Ser Ala Ile Ile His Arg Leu Arg Arg Arg Gly 245 250 255		
Arg Ser Gly Leu Gln Pro Arg Val Gly Lys Ser Thr Ala Ile Leu Leu 260 265 270		
Gly Ile Thr Thr Leu Phe Thr Leu Leu Trp Ala Pro Arg Val Phe Val 275 280 285		
Met Leu Tyr His Met Tyr Val Ala Pro Val His Arg Asp Trp Arg Val 290 295 300		
His Leu Ala Leu Asp Val Ala Asn Met Val Ala Met Leu His Thr Ala 305 310 315 320		
Ala Asn Phe Gly Leu Tyr Cys Phe Val Ser Lys Thr Phe Arg Ala Thr 325 330 335		
Val Arg Gln Val Ile His Asp Ala Tyr Leu Pro Cys Thr Leu Ala Ser 340 345 350		
Gln Pro Glu Gly Met Ala Ala Lys Pro Val Met Glu Pro Pro Gly Leu 355 360 365		

Pro Thr Gly Ala Glu Val
370

<210> 7
<211> 1086
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 7
atgtccactg aatgcgcgcg ggcagcgggc gacgcgccct tgcgcagcct ggagcaagcc 60
aaccgcaccc gctttccctt cttctccgac gtcaaggcgc accaccggct ggtgctggcc 120
gcggtggaga caacgtgct ggtgctcctc tttgcagtgt cgtgctggg caacgtgtgc 180
gccctggtgc tggtagcgcg ccgacgacgc cgcggcgcga ctgcctgcct ggtactcaac 240
ctcttctgcg cggacctgct cttcatcagc gctatccctc tggtagctggc cgtgcgctgg 300
actgaggcct ggctgctggg ccccgttgcc tgccacctgc tcttctacgt gatgacctg 360
agcggcagcg tcaccatcct cacgtggcc gcggtcagcc tggagcgcgt ggtgtgcatc 420
gtgcacctgc agcgcggcgt gcggggtcct gggcggcggg cgcgggcagt gctgctggcg 480
ctcatatggg gctattcggc ggtcgcgcgt ctgcctctat gcgtcttctt ccgagtcgtc 540
ccgcaacggc tccccggcgc cgaccaggaa atttcgattt gcacactgat ttggcccacc 600
attcctggag agatctcgtg gtagtctctt tttgttactt tgaacttctt ggtgccagga 660
ctgggtcattg tgatcagtta ctccaaaatt ttacagatca caaaggcatc aaggaagagg 720
ctcacggtaa gcctggccta ctcgagagc caccagatcc gcgtgtccca gcaggacttc 780
cggtctcttc gcacctctt cctcctcatg gtctccttct tcatcatgtg gagccccatc 840
atcatcacca tcctcctcat cctgatccag aacttcaagc aagacctggt catctggccg 900
tcctcttctt tctgggtggt ggccttcaca tttgctaatt cagccctaaa ccccatcctc 960
tacaacatga cactgtgcag gaatgagtgg aagaaaattt tttgctgctt ctggttccca 1020
gaaaaggagg ccattttaac agacacatct gtcaaaagaa atgacttgct gattatttct 1080
ggctaa 1086

<210> 8
<211> 361
<212> PRT
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 8

Met Ser Thr Glu Cys Ala Arg Ala Ala Gly Asp Ala Pro Leu Arg Ser
1 5 10 15

Leu Glu Gln Ala Asn Arg Thr Arg Phe Pro Phe Phe Ser Asp Val Lys
 20 25 30
 Gly Asp His Arg Leu Val Leu Ala Ala Val Glu Thr Thr Val Leu Val
 35 40 45
 Leu Ile Phe Ala Val Ser Leu Leu Gly Asn Val Cys Ala Leu Val Leu
 50 55 60
 Val Ala Arg Arg Arg Arg Arg Gly Ala Thr Ala Cys Leu Val Leu Asn
 65 70 75 80
 Leu Phe Cys Ala Asp Leu Leu Phe Ile Ser Ala Ile Pro Leu Val Leu
 85 90 95
 Ala Val Arg Trp Thr Glu Ala Trp Leu Leu Gly Pro Val Ala Cys His
 100 105 110
 Leu Leu Phe Tyr Val Met Thr Leu Ser Gly Ser Val Thr Ile Leu Thr
 115 120 125
 Leu Ala Ala Val Ser Leu Glu Arg Met Val Cys Ile Val His Leu Gln
 130 135 140
 Arg Gly Val Arg Gly Pro Gly Arg Arg Ala Arg Ala Val Leu Leu Ala
 145 150 155 160
 Leu Ile Trp Gly Tyr Ser Ala Val Ala Ala Leu Pro Leu Cys Val Phe
 165 170 175
 Phe Arg Val Val Pro Gln Arg Leu Pro Gly Ala Asp Gln Glu Ile Ser
 180 185 190
 Ile Cys Thr Leu Ile Trp Pro Thr Ile Pro Gly Glu Ile Ser Trp Asp
 195 200 205
 Val Ser Phe Val Thr Leu Asn Phe Leu Val Pro Gly Leu Val Ile Val
 210 215 220
 Ile Ser Tyr Ser Lys Ile Leu Gln Ile Thr Lys Ala Ser Arg Lys Arg
 225 230 235 240
 Leu Thr Val Ser Leu Ala Tyr Ser Glu Ser His Gln Ile Arg Val Ser
 245 250 255
 Gln Gln Asp Phe Arg Leu Phe Arg Thr Leu Phe Leu Leu Met Val Ser
 260 265 270
 Phe Phe Ile Met Trp Ser Pro Ile Ile Ile Thr Ile Leu Leu Ile Leu

275

280

285

Ile Gln Asn Phe Lys Gln Asp Leu Val Ile Trp Pro Ser Leu Phe Phe
 290 295 300

Trp Val Val Ala Phe Thr Phe Ala Asn Ser Ala Leu Asn Pro Ile Leu
 305 310 315 320

Tyr Asn Met Thr Leu Cys Arg Asn Glu Trp Lys Lys Ile Phe Cys Cys
 325 330 335

Phe Trp Phe Pro Glu Lys Gly Ala Ile Leu Thr Asp Thr Ser Val Lys
 340 345 350

Arg Asn Asp Leu Ser Ile Ile Ser Gly
 355 360

<210> 9
 <211> 1038
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 9
 atgagcagca attcatccct gctggtggct gtgcagctgt gctacgcgaa cgtgaatggg 60
 tcctgtgtga aaatcccctt ctgcgcggga tcccggtga ttctgtacat agtgtttggc 120
 tttggggctg tgctggctgt gtttggaac ctctggtga tgatttcaat cctccatttc 180
 aagcagctgc actctccgac caattttctc gttgcctctc tggcctgcgc tgatttcttg 240
 gtgggtgtga ctgtgatgcc cttcagcatg gtcaggacgg tggagagctg ctggtatttt 300
 gggaggagt ttgtacttt ccacacctgc tgtgatgtgg cattttgtta ctcttctctc 360
 tttcacttgt gttcatctc catcgacagg tacattgcgg ttactgacct cctggtctat 420
 cctaccaagt tcaccgtatc tgtgtcagga atttgcatca gcgtgtcctg gatcctgccc 480
 ctcatgtaca gcggtgctgt gttctacaca ggtgtctatg acgatgggct ggaggaatta 540
 tctgatgccc taaactgtat aggaggttgt cagaccgttg taaatcaaaa ctgggtgttg 600
 acagattttc tatccttctt tatacctacc tttattatga taattctgta tggtaacata 660
 tttcttgtgg ctagacgaca ggcgaaaaag atagaaaata ctggtagcaa gacagaatca 720
 tcctcagaga gttacaaagc cagagtggcc aggagagaga gaaaagcagc taaaaccctg 780
 ggggtcacag tggtagcatt tatgatttca tggttaccat atagcattga ttcattaatt 840
 gatgccctta tgggctttat aaccctgcc tgtatttatg agatttgctg ttggtgtgct 900
 tattataact cagccatgaa tcctttgatt tatgctttat tttacccatg gtttaggaaa 960
 gcaataaaaag ttattgtaac tggtcagggt ttaaagaaca gttcagcaac catgaatttg 1020

ttttctgaac atatataa

1038

<210> 10
<211> 345
<212> PRT
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 10

Met Ser Ser Asn Ser Ser Leu Leu Val Ala Val Gln Leu Cys Tyr Ala
1 5 10 15

Asn Val Asn Gly Ser Cys Val Lys Ile Pro Phe Ser Pro Gly Ser Arg
20 25 30

Val Ile Leu Tyr Ile Val Phe Gly Phe Gly Ala Val Leu Ala Val Phe
35 40 45

Gly Asn Leu Leu Val Met Ile Ser Ile Leu His Phe Lys Gln Leu His
50 55 60

Ser Pro Thr Asn Phe Leu Val Ala Ser Leu Ala Cys Ala Asp Phe Leu
65 70 75 80

Val Gly Val Thr Val Met Pro Phe Ser Met Val Arg Thr Val Glu Ser
85 90 95

Cys Trp Tyr Phe Gly Arg Ser Phe Cys Thr Phe His Thr Cys Cys Asp
100 105 110

Val Ala Phe Cys Tyr Ser Ser Leu Phe His Leu Cys Phe Ile Ser Ile
115 120 125

Asp Arg Tyr Ile Ala Val Thr Asp Pro Leu Val Tyr Pro Thr Lys Phe
130 135 140

Thr Val Ser Val Ser Gly Ile Cys Ile Ser Val Ser Trp Ile Leu Pro
145 150 155 160

Leu Met Tyr Ser Gly Ala Val Phe Tyr Thr Gly Val Tyr Asp Asp Gly
165 170 175

Leu Glu Glu Leu Ser Asp Ala Leu Asn Cys Ile Gly Gly Cys Gln Thr
180 185 190

Val Val Asn Gln Asn Trp Val Leu Thr Asp Phe Leu Ser Phe Phe Ile
195 200 205

Pro Thr Phe Ile Met Ile Ile Leu Tyr Gly Asn Ile Phe Leu Val Ala

210 215 220
 Arg Arg Gln Ala Lys Lys Ile Glu Asn Thr Gly Ser Lys Thr Glu Ser
 225 230 235 240
 Ser Ser Glu Ser Tyr Lys Ala Arg Val Ala Arg Arg Glu Arg Lys Ala
 245 250 255
 Ala Lys Thr Leu Gly Val Thr Val Val Ala Phe Met Ile Ser Trp Leu
 260 265 270
 Pro Tyr Ser Ile Asp Ser Leu Ile Asp Ala Phe Met Gly Phe Ile Thr
 275 280 285
 Pro Ala Cys Ile Tyr Glu Ile Cys Cys Trp Cys Ala Tyr Tyr Asn Ser
 290 295 300
 Ala Met Asn Pro Leu Ile Tyr Ala Leu Phe Tyr Pro Trp Phe Arg Lys
 305 310 315 320
 Ala Ile Lys Val Ile Val Thr Gly Gln Val Leu Lys Asn Ser Ser Ala
 325 330 335
 Thr Met Asn Leu Phe Ser Glu His Ile
 340 345

<210> 11
 <211> 1020
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 11
 atgatgccct ttgccacaa tataattaat atttcctgtg tgaaaaacaa ctggtcaaatt 60
 gatgtccgtg ctccctgta cagttaaatg gtgctcataa ttctgaccac actcgttggc 120
 aatctgatag ttattgtttc tatatcacac ttcaaacaac ttcatacccc aacaaattgg 180
 ctcatcatt ccatggccac tgtggacttt cttctggggt gtctggtcat gccttacagt 240
 atgggtgagat ctgctgagca ctgttggtat ttgggagaag tcttctgtaa aattcacaca 300
 agcaccgaca ttatgctgag ctgagcctcc attttccatt tgtctttcat ctccattgac 360
 cgctactatg ctgtgtgtga tccactgaga tataaagcca agatgaatat cttggttatt 420
 tgtgtgatga tcttcattag ttggagtgtc cctgctgttt ttgcatttgg aatgatcttt 480
 ctggagctaa acttcaaagg cgctgaagag atatattaca aacatgttca ctgcagagga 540
 ggttgctctg tcttctttag caaaatatct ggggtactga cctttatgac ttctttttat 600
 atacctggat ctattatgtt atgtgtctat tacagaatat atcttatcgc taaagaacag 660

gcaagattaa ttagtgatgc caatcagaag ctccaaattg gattggaaat gaaaaatgga 720
 atttcacaaa gcaaagaaag gaaagctgtg aagacattgg ggattgtgat gggagttttc 780
 ctaatatgct ggtgcccttt ctttatctgt acagtcattgg acccttttct tcaactacatt 840
 attccaccta ctttgaatga tgtattgatt tggtttggct acttgaactc tacatttaat 900
 ccaatggttt atgcattttt ctatccttgg tttagaaaag cactgaagat gatgctgttt 960
 ggtaaaattt tccaaaaaga ttcattccagg tgtaaattat ttttgaatt gagttcatag 1020

<210> 12
 <211> 339
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 12

Met Met Pro Phe Cys His Asn Ile Ile Asn Ile Ser Cys Val Lys Asn
 1 5 10 15

Asn Trp Ser Asn Asp Val Arg Ala Ser Leu Tyr Ser Leu Met Val Leu
 20 25 30

Ile Ile Leu Thr Thr Leu Val Gly Asn Leu Ile Val Ile Val Ser Ile
 35 40 45

Ser His Phe Lys Gln Leu His Thr Pro Thr Asn Trp Leu Ile His Ser
 50 55 60

Met Ala Thr Val Asp Phe Leu Leu Gly Cys Leu Val Met Pro Tyr Ser
 65 70 75 80

Met Val Arg Ser Ala Glu His Cys Trp Tyr Phe Gly Glu Val Phe Cys
 85 90 95

Lys Ile His Thr Ser Thr Asp Ile Met Leu Ser Ser Ala Ser Ile Phe
 100 105 110

His Leu Ser Phe Ile Ser Ile Asp Arg Tyr Tyr Ala Val Cys Asp Pro
 115 120 125

Leu Arg Tyr Lys Ala Lys Met Asn Ile Leu Val Ile Cys Val Met Ile
 130 135 140

Phe Ile Ser Trp Ser Val Pro Ala Val Phe Ala Phe Gly Met Ile Phe
 145 150 155 160

Leu Glu Leu Asn Phe Lys Gly Ala Glu Glu Ile Tyr Tyr Lys His Val
 165 170 175

His Cys Arg Gly Gly Cys Ser Val Phe Phe Ser Lys Ile Ser Gly Val
180 185 190

Leu Thr Phe Met Thr Ser Phe Tyr Ile Pro Gly Ser Ile Met Leu Cys
195 200 205

Val Tyr Tyr Arg Ile Tyr Leu Ile Ala Lys Glu Gln Ala Arg Leu Ile
210 215 220

Ser Asp Ala Asn Gln Lys Leu Gln Ile Gly Leu Glu Met Lys Asn Gly
225 230 235 240

Ile Ser Gln Ser Lys Glu Arg Lys Ala Val Lys Thr Leu Gly Ile Val
245 250 255

Met Gly Val Phe Leu Ile Cys Trp Cys Pro Phe Phe Ile Cys Thr Val
260 265 270

Met Asp Pro Phe Leu His Tyr Ile Ile Pro Pro Thr Leu Asn Asp Val
275 280 285

Leu Ile Trp Phe Gly Tyr Leu Asn Ser Thr Phe Asn Pro Met Val Tyr
290 295 300

Ala Phe Phe Tyr Pro Trp Phe Arg Lys Ala Leu Lys Met Met Leu Phe
305 310 315 320

Gly Lys Ile Phe Gln Lys Asp Ser Ser Arg Cys Lys Leu Phe Leu Glu
325 330 335

Leu Ser Ser

<210> 13
<211> 1029
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 13
atgaccagca atttttccca acctgttgtg cagctttgct atgaggatgt gaatggatct 60
tgtattgaaa ctccctattc tcctgggtcc cggttaattc tgtacacggc gtttagcttt 120
gggtctttgc tggtgtatt tggaatctc ttagtaatga cttctgttct tcattttaag 180
cagctgcact ctccaaccaa ttttctcatt gcctctctgg cctgtgctga cttcttggtg 240
gggtgtgactg tgatgctttt cagcatggtc aggacggtgg agagctgctg gtattttgga 300
gccaaatttt gtactcttca cagttgctgt gatgtggcat tttgttactc tctgtcctc 360

cacttgtgct tcactctgcat cgacaggtac atttgtggtta ctgatcccct ggtctatgct 420
 accaagttca ccgtgtctgt gtcggaatt tgcacagcg tgcctggat tctgcctctc 480
 acgtacagcg gtgctgtgtt ctacacaggt gtcaatgatg atgggctgga ggaattagta 540
 agtgccttca actgcgtagg tggctgtcaa attattgtaa gtcaaggctg ggtgttgata 600
 gattttctgt tattcttcat acctaccctt gttatgataa ttctttacag taagattttt 660
 cttatagcta aacaacaagc tataaaaatt gaaactacta gtagcaaagt agaatcatcc 720
 tcagagagtt ataaaatcag agtggccaag agagagagga aagcagctaa aaccctgggg 780
 gtcacggtac tagcatttgt tatttcatgg ttaccgtata cagttgatat attaattgat 840
 gcctttatgg gcttcctgac cctgcctat atctatgaaa ttgctgttg gagtgcctat 900
 tataactcag ccatgaatcc tttgatttat gctctatttt atccttggtt taggaaagcc 960
 ataaaactta ttttaagtgg agatgtttta aaggctagtt catcaaccat tagtttattt 1020
 ttagaataa 1029

<210> 14
 <211> 342
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 14

Met Thr Ser Asn Phe Ser Gln Pro Val Val Gln Leu Cys Tyr Glu Asp
 1 5 10 15

Val Asn Gly Ser Cys Ile Glu Thr Pro Tyr Ser Pro Gly Ser Arg Val
 20 25 30

Ile Leu Tyr Thr Ala Phe Ser Phe Gly Ser Leu Leu Ala Val Phe Gly
 35 40 45

Asn Leu Leu Val Met Thr Ser Val Leu His Phe Lys Gln Leu His Ser
 50 55 60

Pro Thr Asn Phe Leu Ile Ala Ser Leu Ala Cys Ala Asp Phe Leu Val
 65 70 75 80

Gly Val Thr Val Met Leu Phe Ser Met Val Arg Thr Val Glu Ser Cys
 85 90 95

Trp Tyr Phe Gly Ala Lys Phe Cys Thr Leu His Ser Cys Cys Asp Val
 100 105 110

Ala Phe Cys Tyr Ser Ser Val Leu His Leu Cys Phe Ile Cys Ile Asp
 115 120 125

Arg Tyr Ile Val Val Thr Asp Pro Leu Val Tyr Ala Thr Lys Phe Thr
 130 135 140
 Val Ser Val Ser Gly Ile Cys Ile Ser Val Ser Trp Ile Leu Pro Leu
 145 150 155 160
 Thr Tyr Ser Gly Ala Val Phe Tyr Thr Gly Val Asn Asp Asp Gly Leu
 165 170 175
 Glu Glu Leu Val Ser Ala Leu Asn Cys Val Gly Gly Cys Gln Ile Ile
 180 185 190
 Val Ser Gln Gly Trp Val Leu Ile Asp Phe Leu Leu Phe Phe Ile Pro
 195 200 205
 Thr Leu Val Met Ile Ile Leu Tyr Ser Lys Ile Phe Leu Ile Ala Lys
 210 215 220
 Gln Gln Ala Ile Lys Ile Glu Thr Thr Ser Ser Lys Val Glu Ser Ser
 225 230 235 240
 Ser Glu Ser Tyr Lys Ile Arg Val Ala Lys Arg Glu Arg Lys Ala Ala
 245 250 255
 Lys Thr Leu Gly Val Thr Val Leu Ala Phe Val Ile Ser Trp Leu Pro
 260 265 270
 Tyr Thr Val Asp Ile Leu Ile Asp Ala Phe Met Gly Phe Leu Thr Pro
 275 280 285
 Ala Tyr Ile Tyr Glu Ile Cys Cys Trp Ser Ala Tyr Tyr Asn Ser Ala
 290 295 300
 Met Asn Pro Leu Ile Tyr Ala Leu Phe Tyr Pro Trp Phe Arg Lys Ala
 305 310 315 320
 Ile Lys Leu Ile Leu Ser Gly Asp Val Leu Lys Ala Ser Ser Ser Thr
 325 330 335
 Ile Ser Leu Phe Leu Glu
 340

<210> 15
 <211> 1062
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Novel Sequence
 <400> 15

atggagcaca cgcacgcccc cctcgcagcc aacagctcgc tgtcttgggtg gtcccccggc 60
 tcggcctcgc gcttggggttt cgtgcccgtg gtctactaca gcctcttgct gtgcctcggg 120
 ttaccagcaa atatcttgac agtgatcatc ctctcccagc tgggtggcaag aagacagaag 180
 tcctcctaca actatctctt ggcactcgtc gctgccgaca tcttggtcct ctttttcata 240
 gtgtttgtgg acttccctgtt ggaagatttc atcttgaaca tgcagatgcc tcagggtcccc 300
 gacaagatca tagaagtgtc ggaattctca tccatccaca cctccatatg gattactgta 360
 ccgttaacca ttgacaggta tatcgtgtc tgccaccgc tcaagtacca cacggtctca 420
 taccagccc gcacccggaa agtcattgta agtgtttaca tcaactgctt cctgaccagc 480
 atcccctatt actggtggcc caacatctgg actgaagact acatcagcac ctctgtgcat 540
 cagctcctca tctggatcca ctgcttcacc gtctacctgg tgccctgctc catcttcttc 600
 atcttgaact caatcattgt gtacaagctc aggaggaaga gcaattttcg tctccgtggc 660
 tactccacgg ggaagaccac cgccatcttg ttcaccatta cctccatctt tgccacactt 720
 tgggcccccc gcatcatcat gattctttac cacctctatg gggcgcccat ccagaaccgc 780
 tggctggtgc acatcatgtc cgacattgcc aacatgctag cccttctgaa cacagccatc 840
 aacttcttcc tctactgctt catcagcaag cggttccgca ccatggcagc cgccacgctc 900
 aaggctttct tcaagtgcc gaagcaacct gtacagtctt acaccaatca taacttttcc 960
 ataacaagta gccctggat ctgcgggca aactcacact gcatcaagat gctggtgtac 1020
 cagtatgaca aaaatggaaa acctataaaa gtatccccgt ga 1062

<210> 16
 <211> 353
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 16

Met Glu His Thr His Ala His Leu Ala Ala Asn Ser Ser Leu Ser Trp
1 5 10 15

Trp Ser Pro Gly Ser Ala Cys Gly Leu Gly Phe Val Pro Val Val Tyr
20 25 30

Tyr Ser Leu Leu Leu Cys Leu Gly Leu Pro Ala Asn Ile Leu Thr Val
35 40 45

Ile Ile Leu Ser Gln Leu Val Ala Arg Arg Gln Lys Ser Ser Tyr Asn
50 55 60

Tyr Leu Leu Ala Leu Ala Ala Ala Asp Ile Leu Val Leu Phe Phe Ile
65 70 75 80

Val Phe Val Asp Phe Leu Leu Glu Asp Phe Ile Leu Asn Met Gln Met
85 90 95
Pro Gln Val Pro Asp Lys Ile Ile Glu Val Leu Glu Phe Ser Ser Ile
100 105 110
His Thr Ser Ile Trp Ile Thr Val Pro Leu Thr Ile Asp Arg Tyr Ile
115 120 125
Ala Val Cys His Pro Leu Lys Tyr His Thr Val Ser Tyr Pro Ala Arg
130 135 140
Thr Arg Lys Val Ile Val Ser Val Tyr Ile Thr Cys Phe Leu Thr Ser
145 150 155 160
Ile Pro Tyr Tyr Trp Trp Pro Asn Ile Trp Thr Glu Asp Tyr Ile Ser
165 170 175
Thr Ser Val His His Val Leu Ile Trp Ile His Cys Phe Thr Val Tyr
180 185 190
Leu Val Pro Cys Ser Ile Phe Phe Ile Leu Asn Ser Ile Ile Val Tyr
195 200 205
Lys Leu Arg Arg Lys Ser Asn Phe Arg Leu Arg Gly Tyr Ser Thr Gly
210 215 220
Lys Thr Thr Ala Ile Leu Phe Thr Ile Thr Ser Ile Phe Ala Thr Leu
225 230 235 240
Trp Ala Pro Arg Ile Ile Met Ile Leu Tyr His Leu Tyr Gly Ala Pro
245 250 255
Ile Gln Asn Arg Trp Leu Val His Ile Met Ser Asp Ile Ala Asn Met
260 265 270
Leu Ala Leu Leu Asn Thr Ala Ile Asn Phe Phe Leu Tyr Cys Phe Ile
275 280 285
Ser Lys Arg Phe Arg Thr Met Ala Ala Ala Thr Leu Lys Ala Phe Phe
290 295 300
Lys Cys Gln Lys Gln Pro Val Gln Phe Tyr Thr Asn His Asn Phe Ser
305 310 315 320
Ile Thr Ser Ser Pro Trp Ile Ser Pro Ala Asn Ser His Cys Ile Lys
325 330 335
Met Leu Val Tyr Gln Tyr Asp Lys Asn Gly Lys Pro Ile Lys Val Ser

Pro

<210> 17
 <211> 969
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 17
 atggatccaa ccgtcccagt cttcgggtaca aaactgacac caatcaacgg acgtgaggag 60
 actccttgct acaatcagac cctgagcttc acgggtgctga cgtgcatcat ttcccttgtc 120
 ggactgacag gaaacgcggt agtgctctgg ctctctgggt accgcatgcg caggaacgct 180
 gtctccatct acatcctcaa cctggccgca gcagacttcc tcttcctcag cttccagatt 240
 atacgttcgc cattacgcct catcaatata agccatctca tccgcaaaat cctcgtttct 300
 gtgatgacct ttccctactt tacaggcctg agtatgctga gcgccatcag caccgagcgc 360
 tgctgtctg ttctgtggcc catctgggtac cgctgccgcc gccccacaca cctgtcagcg 420
 gtcgtgtgtg tcctgtctctg gggcctgtcc ctgctgttta gtatgctgga gtggagggtc 480
 tgtgacttcc tgtttagtgg tgctgattct agttggtgtg aaacgtcaga ttcatccca 540
 gtcgcgtggc tgattttttt atgtgtgggt ctctgtgttt ccagcctggt cctgctggtc 600
 aggatcctct gtggatcccg gaagatgccg ctgaccaggc tgtacgtgac catcctgctc 660
 acagtgtctg tcttcctcct ctgcggcctg cccttcggca ttctgggggc cctaatttac 720
 aggatgcacc tgaatttgga agtcttatat tgtcatgttt atctggtttg catgtccctg 780
 tcctctctaa acagtagtgc caaccccatc atttacttct tcgtgggctc cttaggcag 840
 cgtcaaaaata ggcagaacct gaagctgggt ctccagaggg ctctgcagga caagcctgag 900
 gtggataaag gtgaaggcca gcttcctgag gaaagcctgg agctgtcggg aagcagattg 960
 gggccatga 969

<210> 18
 <211> 322
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 18

Met Asp Pro Thr Val Pro Val Phe Gly Thr Lys Leu Thr Pro Ile Asn
 1 5 10 15

Gly Arg Glu Glu Thr Pro Cys Tyr Asn Gln Thr Leu Ser Phe Thr Val

20	25	30
Leu Thr Cys Ile Ile Ser Leu Val Gly Leu Thr Gly Asn Ala Val Val 35 40 45		
Leu Trp Leu Leu Gly Tyr Arg Met Arg Arg Asn Ala Val Ser Ile Tyr 50 55 60		
Ile Leu Asn Leu Ala Ala Ala Asp Phe Leu Phe Leu Ser Phe Gln Ile 65 70 75 80		
Ile Arg Ser Pro Leu Arg Leu Ile Asn Ile Ser His Leu Ile Arg Lys 85 90 95		
Ile Leu Val Ser Val Met Thr Phe Pro Tyr Phe Thr Gly Leu Ser Met 100 105 110		
Leu Ser Ala Ile Ser Thr Glu Arg Cys Leu Ser Val Leu Trp Pro Ile 115 120 125		
Trp Tyr Arg Cys Arg Arg Pro Thr His Leu Ser Ala Val Val Cys Val 130 135 140		
Leu Leu Trp Gly Leu Ser Leu Leu Phe Ser Met Leu Glu Trp Arg Phe 145 150 155 160		
Cys Asp Phe Leu Phe Ser Gly Ala Asp Ser Ser Trp Cys Glu Thr Ser 165 170 175		
Asp Phe Ile Pro Val Ala Trp Leu Ile Phe Leu Cys Val Val Leu Cys 180 185 190		
Val Ser Ser Leu Val Leu Leu Val Arg Ile Leu Cys Gly Ser Arg Lys 195 200 205		
Met Pro Leu Thr Arg Leu Tyr Val Thr Ile Leu Leu Thr Val Leu Val 210 215 220		
Phe Leu Leu Cys Gly Leu Pro Phe Gly Ile Leu Gly Ala Leu Ile Tyr 225 230 235 240		
Arg Met His Leu Asn Leu Glu Val Leu Tyr Cys His Val Tyr Leu Val 245 250 255		
Cys Met Ser Leu Ser Ser Leu Asn Ser Ser Ala Asn Pro Ile Ile Tyr 260 265 270		
Phe Phe Val Gly Ser Phe Arg Gln Arg Gln Asn Arg Gln Asn Leu Lys 275 280 285		

Leu Val Leu Gln Arg Ala Leu Gln Asp Lys Pro Glu Val Asp Lys Gly
 290 295 300

Glu Gly Gln Leu Pro Glu Glu Ser Leu Glu Leu Ser Gly Ser Arg Leu
 305 310 315 320

Gly Pro

<210> 19
 <211> 969
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 19
 atggattcaa ccatcccagt cttgggtaca gaactgacac caatcaacgg acgtgaggag 60
 actccttgct acaagcagac cctgagcttc acggggctga cgtgcatcgt ttcccttgct 120
 gcgctgacag gaaacgcggt tgtgctctgg ctccctgggt gccgcatgag caggaacgct 180
 gtctccatct acatcctcaa cctggctcgc gccgacttcc tcttccttag cggccacatt 240
 atatgttcgc cgttacgcct catcaatata cgccatccca tctccaaaat cctcagtcct 300
 gtgatgacct ttccctactt tataggccta agcatgctga gcgccatcag caccgagcgc 360
 tgcctgtcca tcctgtggcc catctggtac cactgccgcc gcccagata cctgtcatcg 420
 gtcattgtgt tcctgtctct ggccctgtcc ctgctgcgga gtatcctgga gtggatgttc 480
 tgtgacttcc tgcttagtgg tgctgattct gtttggtgtg aaacgtcaga ttccattaca 540
 atcgcgtggc tggttttttt atgtgtgggt ctctgtgggt ccagcctggt cctgctggtc 600
 aggattctct gtggatcccg gaagatgccg ctgaccaggc tgtacgtgac catcctcctc 660
 acagtgtctg tcttcctcct ctgtggcctg ccctttggca ttcagtgggc cctgttttcc 720
 aggatccacc tggattggaa agtcttattt tgtcatgtgc atctagtctc catcttcctg 780
 tccgctctta acagcagtgc caaccccatc atttacttct tcgtgggctc ctttaggcag 840
 cgtcaaaata ggcagaacct gaagctgggt ctccagaggg ctctgcagga caccgctgag 900
 gtggatgaag gtggaggggt gcttcctcag gaaaccctgg agctgtcggg aagcagattg 960
 gagcagtga 969

<210> 20
 <211> 322
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 20

Met Asp Ser Thr Ile Pro Val Leu Gly Thr Glu Leu Thr Pro Ile Asn
1 5 10 15

Gly Arg Glu Glu Thr Pro Cys Tyr Lys Gln Thr Leu Ser Phe Thr Gly
20 25 30

Leu Thr Cys Ile Val Ser Leu Val Ala Leu Thr Gly Asn Ala Val Val
35 40 45

Leu Trp Leu Leu Gly Cys Arg Met Arg Arg Asn Ala Val Ser Ile Tyr
50 55 60

Ile Leu Asn Leu Val Ala Ala Asp Phe Leu Phe Leu Ser Gly His Ile
65 70 75 80

Ile Cys Ser Pro Leu Arg Leu Ile Asn Ile Arg His Pro Ile Ser Lys
85 90 95

Ile Leu Ser Pro Val Met Thr Phe Pro Tyr Phe Ile Gly Leu Ser Met
100 105 110

Leu Ser Ala Ile Ser Thr Glu Arg Cys Leu Ser Ile Leu Trp Pro Ile
115 120 125

Trp Tyr His Cys Arg Arg Pro Arg Tyr Leu Ser Ser Val Met Cys Val
130 135 140

Leu Leu Trp Ala Leu Ser Leu Leu Arg Ser Ile Leu Glu Trp Met Phe
145 150 155 160

Cys Asp Phe Leu Phe Ser Gly Ala Asp Ser Val Trp Cys Glu Thr Ser
165 170 175

Asp Phe Ile Thr Ile Ala Trp Leu Val Phe Leu Cys Val Val Leu Cys
180 185 190

Gly Ser Ser Leu Val Leu Leu Val Arg Ile Leu Cys Gly Ser Arg Lys
195 200 205

Met Pro Leu Thr Arg Leu Tyr Val Thr Ile Leu Leu Thr Val Leu Val
210 215 220

Phe Leu Leu Cys Gly Leu Pro Phe Gly Ile Gln Trp Ala Leu Phe Ser
225 230 235 240

Arg Ile His Leu Asp Trp Lys Val Leu Phe Cys His Val His Leu Val
245 250 255

Ser Ile Phe Leu Ser Ala Leu Asn Ser Ser Ala Asn Pro Ile Ile Tyr

260 265 270
 Phe Phe Val Gly Ser Phe Arg Gln Arg Gln Asn Arg Gln Asn Leu Lys
 275 280 285
 Leu Val Leu Gln Arg Ala Leu Gln Asp Thr Pro Glu Val Asp Glu Gly
 290 295 300
 Gly Gly Trp Leu Pro Gln Glu Thr Leu Glu Leu Ser Gly Ser Arg Leu
 305 310 315 320

Glu Gln

<210> 21
 <211> 26
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 21
 cagagctctg gtggccacct ctgtcc 26

<210> 22
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 22
 ctgcgtccac cagagtcacg tctcc 25

<210> 23
 <211> 26
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 23
 gtatgcctgg ccacaatacc tccagg 26

<210> 24
 <211> 31
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

<400> 24
 gtttgtggct aacggcacia aacacaattc c 31

<210> 25
 <211> 28
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 25
 ggtaccacaa tgacaatcac cagcgtcc 28

 <210> 26
 <211> 29
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 26
 ggaacgtgag gtacatgtgg atgtgcagc 29

 <210> 27
 <211> 27
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 27
 gcagtgtagc ggtcaaccgt gagcagg 27

 <210> 28
 <211> 32
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 28
 tgagcaggat ggcgatccag actgaggcgt gg 32

 <210> 29
 <211> 25
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 29
 gaggtacagc tggcgtgct gacag 25

 <210> 30
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

 <400> 30
 gtggccatga gccaccctga gctcc 25

 <210> 31
 <211> 24
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 31
 ggaatgtcca ctgaatgcgc gcgg 24

 <210> 32
 <211> 25
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 32
 agctcgccag gtgtgagaaa ctcgg 25

 <210> 33
 <211> 30
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 33
 gcgttatgag cagcaattca tccctgctgg 30

 <210> 34
 <211> 28
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 34
 gtatcctgaa cttcgtctat acaactgc 28

 <210> 35
 <211> 29
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 35
 ccctcaggaa tgatgccctt ttgccacaa 29

<210> 36
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 36
atccatgtgg ttggtgcatg tggttcgt

28

<210> 37
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 37
aaacaacaaa cagcagaacc atgaccagc

29

<210> 38
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 38
acatagagac aagtgacatg tgtgaaccac

30

<210> 39
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 39
ggtatgagac cgtgtggtac ttgagc

26

<210> 40
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 40
gtggcagaca gcgatatacc tgtcaatgg

29

<210> 41
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
 <223> Novel Sequence

 <400> 41
 gcgctcatgg agcacacgca cgccac 27

 <210> 42
 <211> 25
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 42
 gaggcagtag ttgccacacc tatgg 25

 <210> 43
 <211> 29
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 43
 catctggttt gtgttcccag gggcaccag 29

 <210> 44
 <211> 32
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 44
 gacagtgttg ctctcaaagt cccgtctgac tg 32

 <210> 45
 <211> 25
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 45
 ctgtttccag ggtcatcaga ctggg 25

 <210> 46
 <211> 27
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 46
 gcagcattgc tctcaaagtc ctgtctg 27

<210> 47
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 47

Thr Leu Glu Ser Ile Met
1 5

<210> 48
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 48

Glu Tyr Asn Leu Val
1 5

<210> 49
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 49

Asp Cys Gly Leu Phe
1 5

<210> 50
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 50
gatcaagctt ccatggcgtg ctgcctgagc gagg

34

<210> 51
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 51
gatcggatcc ttagaacagg ccgcagtcct tcagggtcag ctgcaggatg gtg

53

<210> 52
 <211> 25
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 52
 gtcctcactg gtggccatgt actcc 25

 <210> 53
 <211> 25
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 53
 ctgctccac cagagtcacg tctcc 25

 <210> 54
 <211> 26
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 54
 cttggatggtt tgggctgccc ttctgc 26

 <210> 55
 <211> 31
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 55
 gtttgtggct aacggcacaa aacacaattc c 31

 <210> 56
 <211> 26
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 56
 ctgctcacgg ttgaccgcta cactgc 26

 <210> 57
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Novel Sequence

 <400> 57
 gtggccatga gccaccctga gctcc 25

 <210> 58
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 58
 cttcttctcc gacgtcaagg 20

 <210> 59
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 59
 cttcttctcc gacgtcaagg 20

 <210> 60
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 60
 cttcttctcc gacgtcaagg 20

 <210> 61
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 61
 caacggtctg acaacctcct 20

 <210> 62
 <211> 21
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Novel Sequence

 <400> 62
 ttgctgtgat gtggcatttt g 21

<210> 63
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 63
caggaagccc ataaaggcat caa 23

<210> 64
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 64
acatcacctg cttcctgacc 20

<210> 65
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 65
ccagcatctt gatgcagtgt 20

<210> 66
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 66
ccatctccaa aatcctcagt c 21

<210> 67
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Novel Sequence

<400> 67
gctgttaaga gcggacagga aa 22